



Estadística Descriptiva y Comparativa

Miguel Arredondo

Estadística: Definición

- La estadística descriptiva sirve al propósito de generar una serie de medidas numéricas tendientes a sintetizar el comportamiento de las variables (o conjunto de datos) que estamos estudiando.
- Así la ***estadística descriptiva*** es un conjunto de métodos y técnicas por medio de las cuales podemos recolectar, organizar, resumir, presentar y analizar datos numéricos relativos a un conjunto de individuos u observaciones.

Estadística Inferencial

- **La estadística descriptiva** no se utiliza como fundamento de una teoría general aplicable a individuos similares que no han sido medidos. Sólo se aplica a la muestra.
- **La estadística inferencial** toma la información de la estadística descriptiva para realizar procesos que permiten, dentro de ciertos rangos de confianza, determinar el comportamiento probable de la población que es representada por la Muestra

El Rol de las Probabilidades en la inferencia estadística



NIVELES DE MEDICIÓN

NIVEL NOMINAL

**Ej: Nacionalidad
Estado civil**

NIVEL ORDINAL

**Ej: NSE
estado nutritivo**

NIVEL INTERVALAR

**Ej: pesos
edad**

Nivel Nominal

- Los números sólo se utilizan con el fin de clasificar los objetos o propiedades o distintas modalidades que asume una variable.
- En este tipo de medición sólo se acepta como relación válida la igualdad o desigualdad entre los números que reflejan en la propiedad medida.

Nivel Ordinal

- Además de realizar la distinción entre las modalidades es posible establecer una **relación de orden o jerarquía**.
- No refiere nada respecto a la modalidad respecto de otra).

Nivel intervalar

- Conserva las propiedades de los niveles anteriores, y agrega la posibilidad de establecer la magnitud en que una modalidad es mayor o menor respecto de otra. (Da información respecto a la distancia entre las modalidades) La diferencia entre las distintas magnitudes es constante.
- El valor cero en este nivel de medición es arbitrario. Esto significa que cero no indica ausencia del atributo, lo que implica que los valores de la variable no son proporcionales entre sí.
- Es un real nivel cuantitativo al cual podemos aplicar las relaciones numéricas de la suma, resta, división y multiplicación.

Medidas de posición

Ubicación respecto al grupo

Medidas de Posición

- Las medidas de posición nos permiten dividir a una distribución en distintas partes. Según el número de partes, éstas divisiones tendrán distintos nombres:
 - Percentiles : Cuando dividimos una distribución en 100 partes.
 - Deciles : Cuando dividimos una distribución en 10 partes.
 - Cuartiles : Cuando dividimos una distribución en 4 partes.
 - Mediana : Cuando dividimos una distribución en 2 partes.

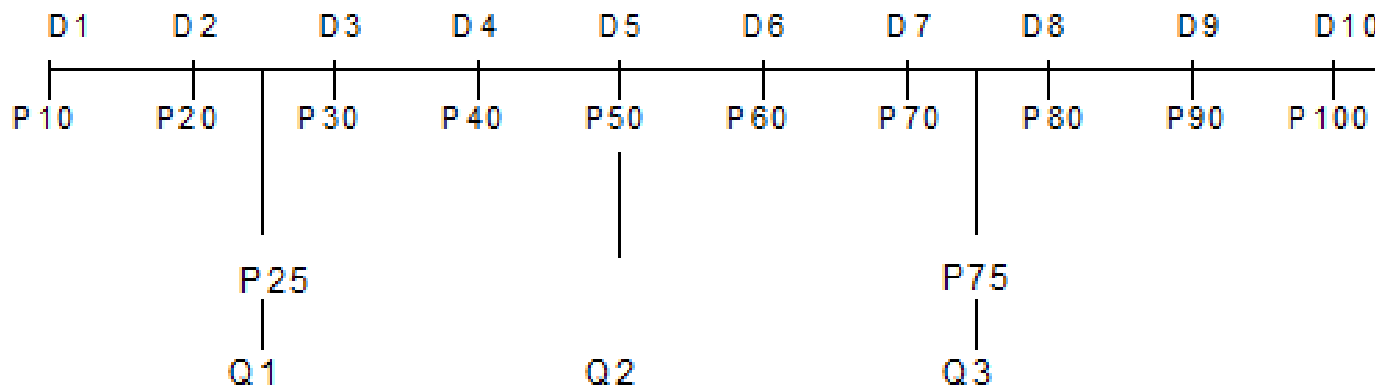
Medidas de posición

- **Requisitos**

- Variables cuantitativas
- Los resultados están ordenados de menor a mayor
- Las medidas de posición son variables discretas

- **Equivalencias**

- Como todas las medidas se refieren al mismo grupo de datos, se pueden hacer equivalentes entre sí



Medidas de Posición

- El percentil 30 (P_{30}), por ejemplo es el valor de la variable bajo el cual se encuentra el 30% de los casos y sobre el cual se encuentra el 70 % de los casos
- El decil 4 (D_4) es el valor de la variable que deja bajo él, el 40 % de los casos y sobre él el 60 %. Esto permite ver las equivalencias entre percentiles y deciles : $(D_4) = (P_{40})$

Cálculo del Percentil

- El cálculo del percentil cuando los datos están “suelos” es directo. Ubicar el puntaje que está en la posición que corresponde al porcentaje del percentil

Ej: 23 34 34 36 37 37 37 42 46 50

$$P_{60}=37$$

- Si los datos están en una tabla de intervalos, la fórmula es la siguiente

$$P_x = L + \left[\frac{\frac{x}{100} * N - F}{f} \right] * i$$

N = Número total de sujetos

L = límite inferior real del intervalo crítico

F = Frecuencia acumulada hasta el intervalo anterior

f = frecuencia relativa en el intervalo crítico

i = Amplitud del intervalo

Medidas de Tendencia Central

El valor más representativo

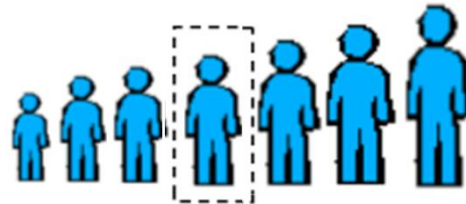
Medidas de Tendencia Central

Son medidas alrededor de las cuales se concentran los datos

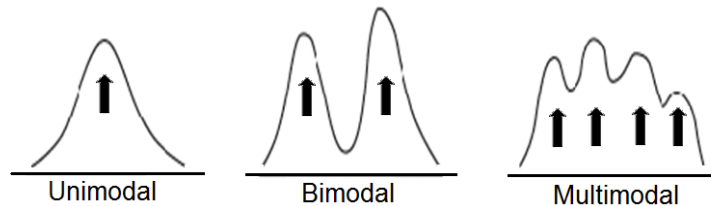
Las tres medidas más usuales de tendencia central son:

Promedio $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

Mediana



Modo



Medidas de Tendencia Central

- Una MTC es un indicador numérico que representa el comportamiento que se considera más representativo de un grupo de valores. Para ello, podemos ocupar distintos criterios:
 - El puntaje que más se repite
 - El que divide al grupo por la mitad
 - El que equipara los puntajes positivos con los negativos

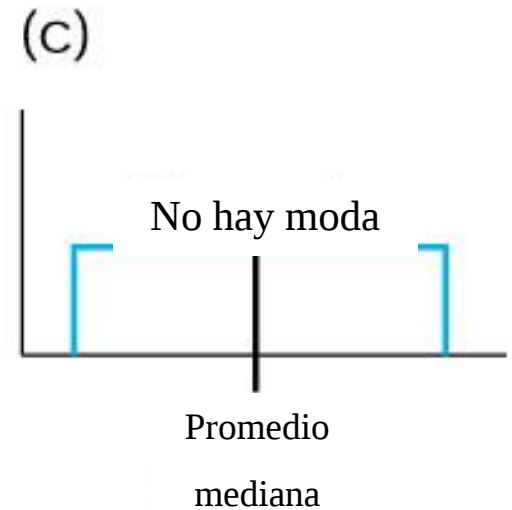
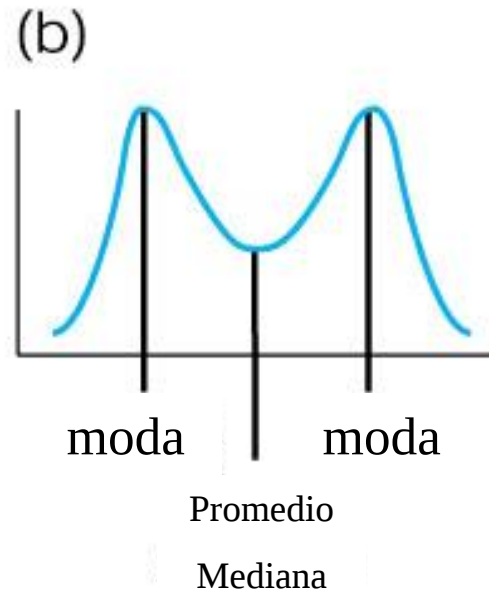
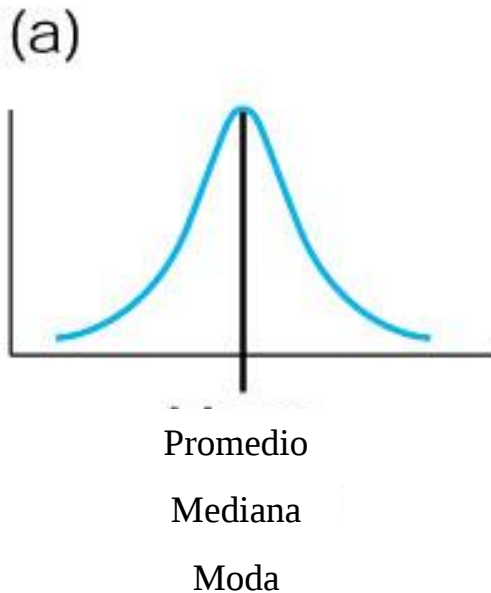
LA MODA

- Representa al puntaje de mayor frecuencia (f) en una [distribución](#) de puntajes.
- Una distribución de puntajes puede ser bimodal, esto es, tener dos modas. Si hay más de dos valores con la mayor frecuencia se dice que no tiene moda.
- En datos no tabulados, es el valor que se repite más.
- En tablas de intervalos, conviene usar una fórmula de nivelación de la moda.

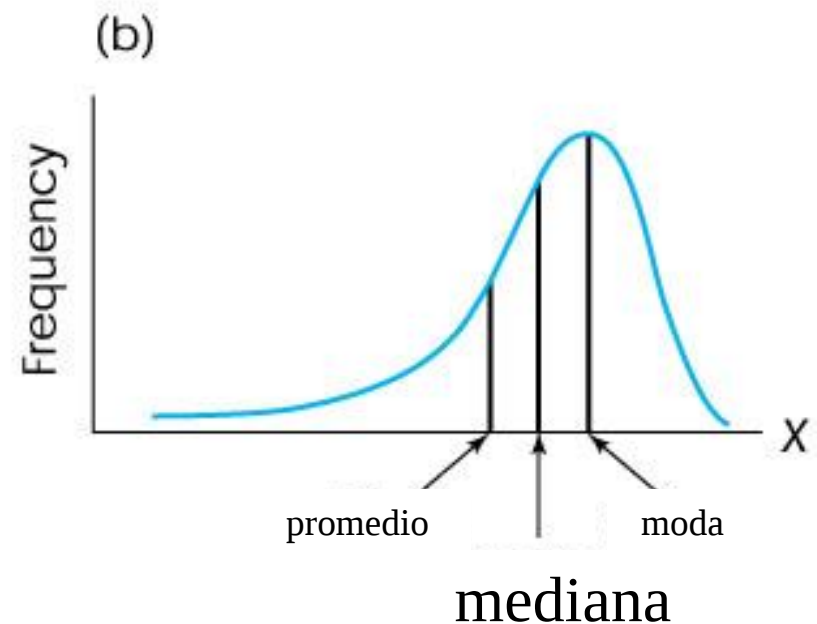
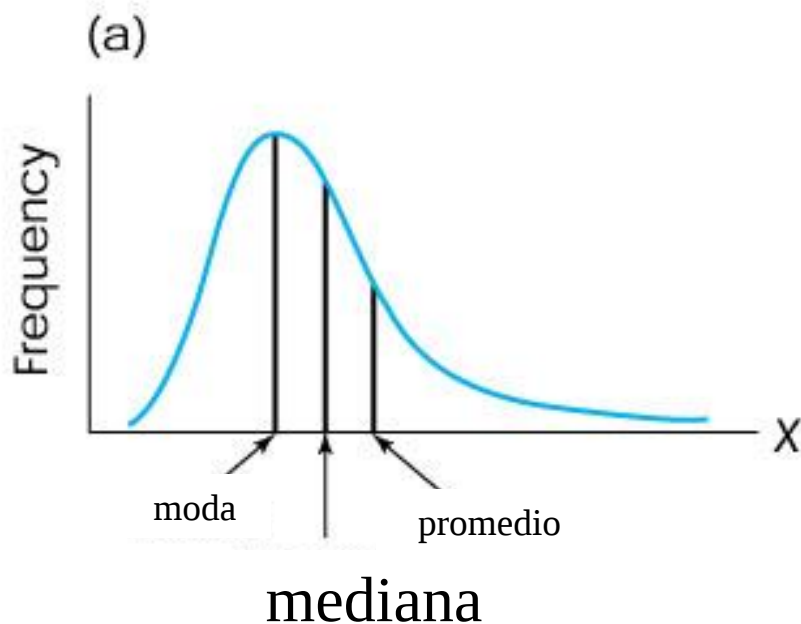
Promedio o Media

- Representa al puntaje que equilibra los valores positivos con los negativos de una distribución.
- Como incluye a toda la distribución, puede ser transformada algebraica-mente
- Soporta variantes: promedio ponderado, agrupaciones de promedios

Medidas de tendencia Central y distribuciones simétricas



Medidas de tendencia Central y distribuciones asimétricas



EL PROMEDIO

Las tres propiedades más importantes del promedio:

- Balancea las diferencias de puntajes
- Refleja las transformaciones hechas a los puntajes
- Puede no ser representativo cuando los valores son extremos.

El promedio: un punto de equilibrio

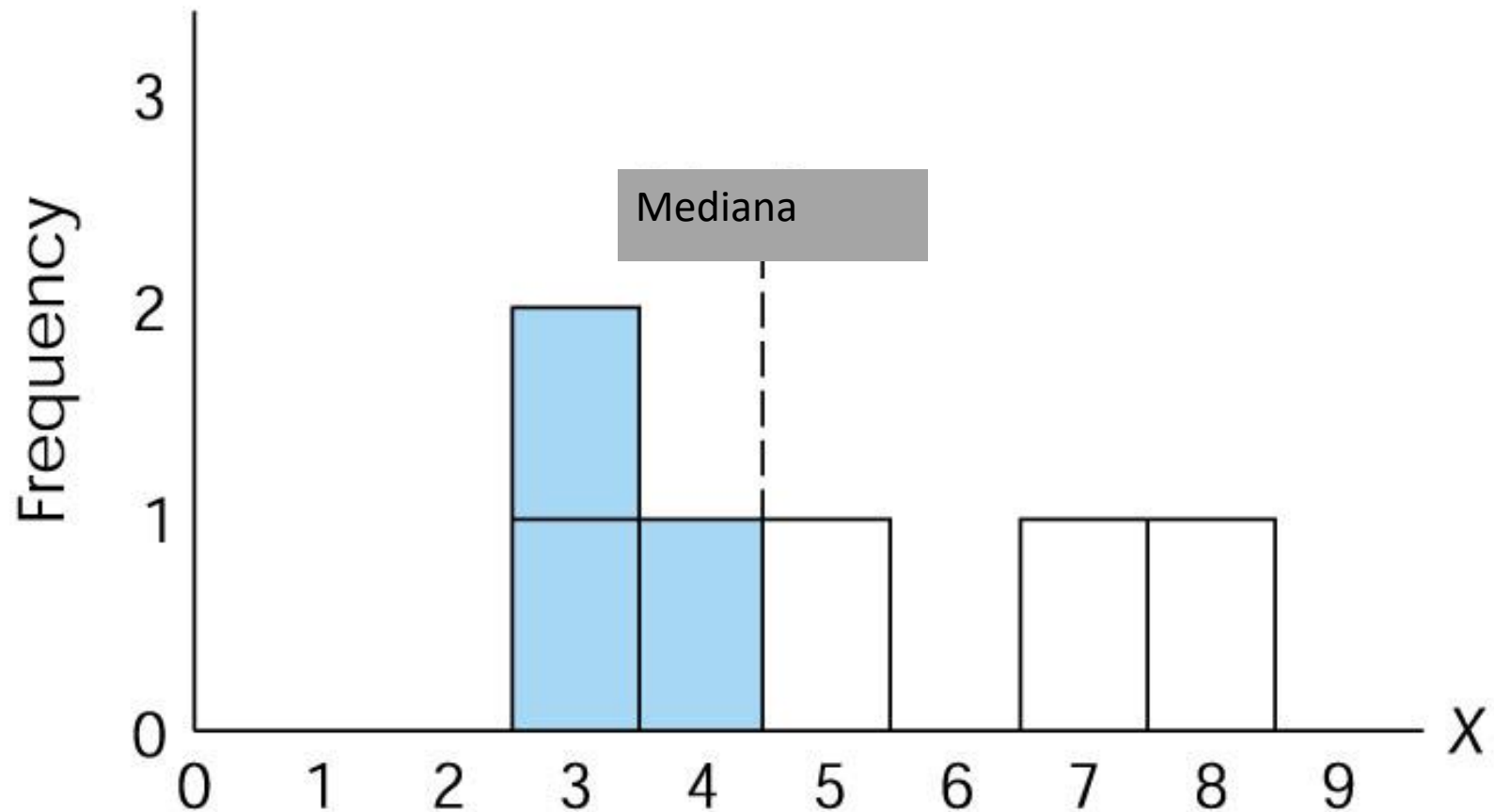
Propiedades del Promedio

1 -Balancea las diferencias de puntajes

			4			
			4			
		3	4	5		
	2	3	4	5	6	
	2	3	4	5	6	
1	2	3	4	5	6	7

PROMEDIO: **4**

LA MEDIANA DIVIDE EL AREA EN DOS



Moda (Mo): Fórmula

- Se aplica sólo en caso de tablas de intervalos:

$$Mo = L + \left[\frac{D_1}{D_1 + D_2} \right] i$$

L = LIR intervalo Modal

D_1 = $f_{\text{modal}} - f_{\text{intervalo anterior}}$

D_2 = $f_{\text{modal}} - f_{\text{intervalo siguiente}}$

I = amplitud del intervalo

Mediana (Md)

- El valor que divide a la distribución por la mitad (equivale a P_{50})

$$Md = L + \left[\frac{N * 0.5 - F}{f} \right] i$$

L = LIR intervalo crítico

N = número total de sujetos

F = Frecuencia acumulada hasta el intervalo anterior al crítico

f = frecuencia del intervalo crítico

i = amplitud del intervalo crítico

Promedio: Fórmula

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

n = número total de puntajes

Σ = suma desde 1 a n

x_i = cada puntaje

El promedio equilibra las diferencias, no las elimina

Equivalencias entre las distintas medidas de posición

- Las **medidas de posición** establecen una relación entre un puntaje y la ubicación que este tiene dentro de la distribución de puntajes de la variable.
- En otras palabras, si una persona, por ejemplo Manuel, tiene 30 puntos en una prueba, y usted quiere saber como le fue a Manuel en relación a su curso, una de las alternativas es determinar qué lugar ocupa el puntaje 30 dentro de todos los puntajes obtenidos en esa prueba.
- El **percentil** responde esa pregunta. Si por ejemplo Manuel está en el percentil 78, quiere decir que con su puntaje está sobre el 78 % de los puntajes de la prueba, o sea, a Manuel le fue bien.

Medidas de Variabilidad

¿Cuán parecidos son los puntajes entre sí?

Medidas de Variabilidad

- Un dato muy importante para analizar las distribuciones de puntajes es saber cuán diferentes o parecidos son los puntajes entre sí
- Esa información nos permite evaluar cuan representativas son las medidas de tendencia central que calculamos para esa misma distribución

Medidas de Dispersión

- Cuantifican la variabilidad de los valores de la distribución respecto al valor central.
- Nos dicen hasta qué punto las medidas de tendencia central son representativas como síntesis de la información.

Las más usadas son:

Varianza

Desviación Estándar

Desviación máxima (Rango)

Desviación cuartil

Error Estándar de la media (EEM)

Medidas de Variabilidad

- ¿Cuáles son?
 - Rango o Amplitud total
 - $R = \text{Máximo} - \text{Mínimo}$
 - Muy fácil de calcular, evalúa la desviación total
 - Rango semi- intercuartil
 - Calcula el punto medio de separación entre cuartil 3 y cuartil 1
 - No es muy usado, difícil de interpretar
 - Varianza – Desviación estándar
 - La más completa

Desviación Estándar

- Considera a todos los valores de la distribución.
- Se utiliza el promedio como punto de referencia para determinar **cómo se desvían** los valores de este punto.
- Mientras más **heterogénea** sea una distribución, mayores serán las desviaciones respecto del promedio

Desviación Estándar

- Lo más sencillo sería sumar las desviaciones de cada punto respecto del promedio.

Puntaje(x)	X-promedio
1	-5
5	-1
6	0
7	1
8	2
9	3

Desviación Estándar

- Sin embargo:
 - Mientras más valores tenga una distribución mayor será la suma de las desviaciones en **valor absoluto** (problema para comparar distribuciones con distinto tamaño).
- Para solucionar esto se calcula **un promedio de las desviaciones**. Se calcula tomando en cuenta las **desviaciones al cuadrado**.

Varianza y Desviación Estándar

- Sin embargo la medida que resulta está expresada en puntajes elevados al cuadrado. Esto se conoce como **VARIANZA** (s^2)

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

- Este indicador nos muestra las variaciones al cuadrado respecto al promedio.
- No se puede interpretar

Desviación estándar

- Para poder interpretar los datos debemos regresarlos a la escala original. Esto se conoce como la **desviación estándar (s)**

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

Puntaje Estándar: Valor Z

- Cualquier escala de puntajes puede transformarse en una nueva serie de puntajes llamados **puntajes estándar (Z)**. Para determinar esta nueva escala de puntajes es necesario definir:
 - 1.1. La distancia entre un puntaje y el promedio
 - 2.2. La desviación estándar que nos indicará cuan heterogéneo u homogéneo es la distribución.
 3. Se dividen ambos valores y se genera esta transformación a un puntaje Z o estándar.

Puntaje Estándar

- Fórmula

$$Z = \frac{X - \bar{X}}{S}$$

$$Z = \frac{X - \bar{X}}{S}$$

Puntaje Z

Un puntaje Z lo que hace es decirnos a cuántas unidades de desviación estándar del promedio está un puntaje determinado, o sea, no contamos en cantidad de puntos, sino en cantidades de desviaciones estándar. Para utilizar el puntaje Z requerimos que la distribución sea normal y Conocer el promedio y la desviación estándar de los puntajes

Distribución de puntajes

- Una distribución de puntajes es el conjunto de valores obtenidos al realizar un experimento estadístico dado. En otras palabras, es el conjunto de resultados de un experimento estadístico.
- Un experimento estadístico es cualquier operación que entrega como producto los valores de una variable (no confundir con un diseño experimental).

DISTRIBUCION NORMAL

Esta distribución de probabilidades es una de las más útiles y usadas de las distribuciones probabilísticas.

- Con mucha frecuencia se encuentran variables aleatorias continuas en la "realidad" (peso, estatura, tiempo, etc que se distribuyen en forma normal.
- Así es de extrema utilidad el uso de la distribución normal teórica para determinar la probabilidad asignada a eventos observados en una variable continua. La utilidad esta en que sirve de apoyo a la toma de decisiones de las hipótesis en término de "aceptarlas" o "rechazarlas".

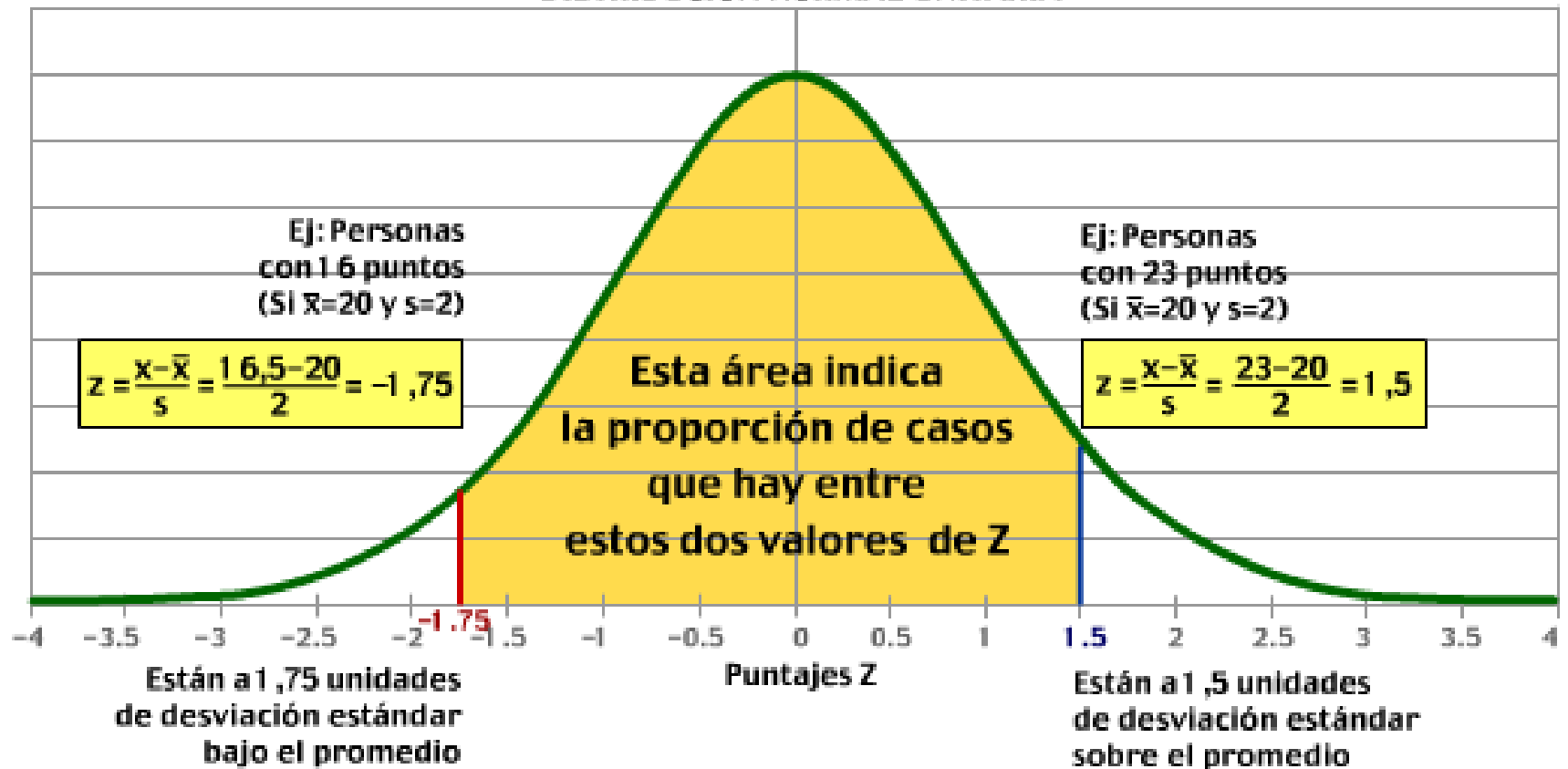
Existen infinitas distribuciones normales, pero la más utilizada es la distribución normal unitaria o estandarizada

Propiedades de la distribución normal unitaria

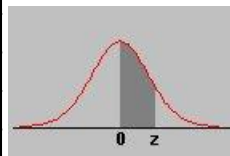
- El promedio de esta distribución es igual a "Cero"
- La desviación estándar de esta distribución es igual a "uno"
- Por tratarse de una distribución de probabilidades, el área total bajo la curva es igual a 1 o 100 %.
- Los valores de las probabilidades asignados a los distintos valores de la v.a. Z (continua) son valores de intervalos.

Interpretando el Puntaje Z

DISTRIBUCION NORMAL UNITARIA

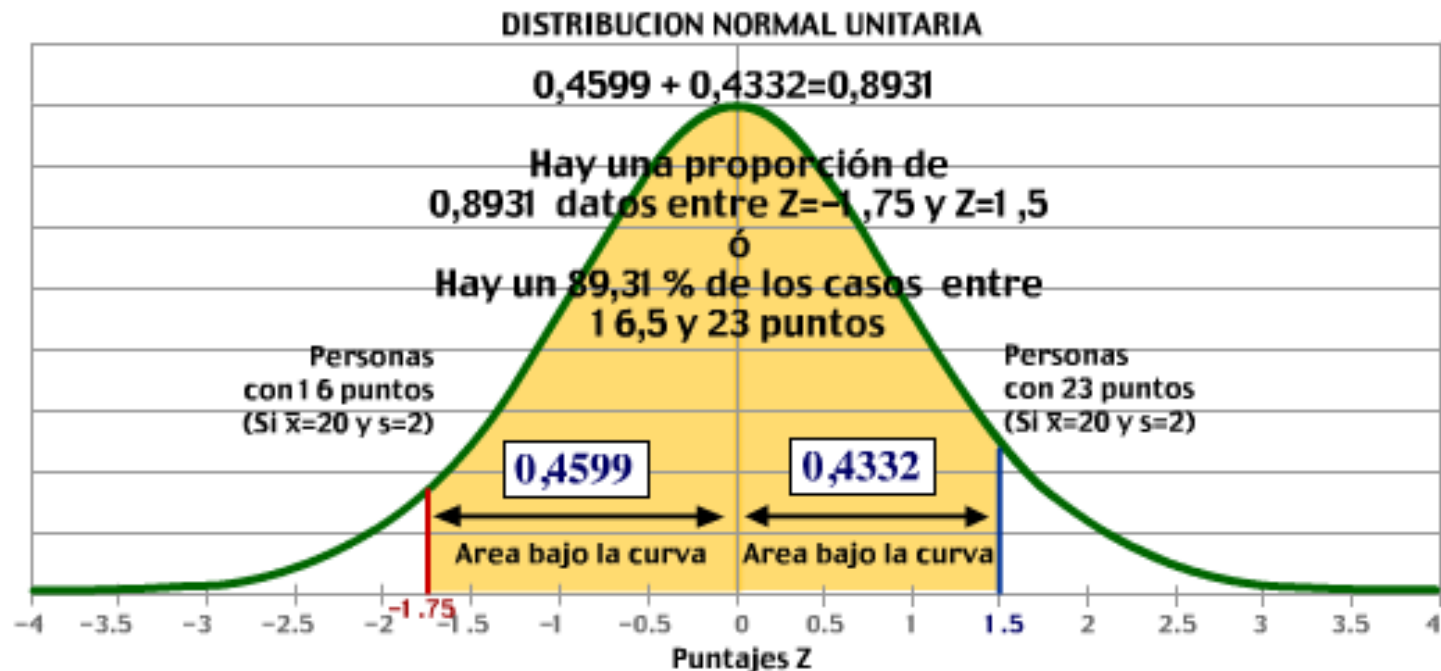


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981

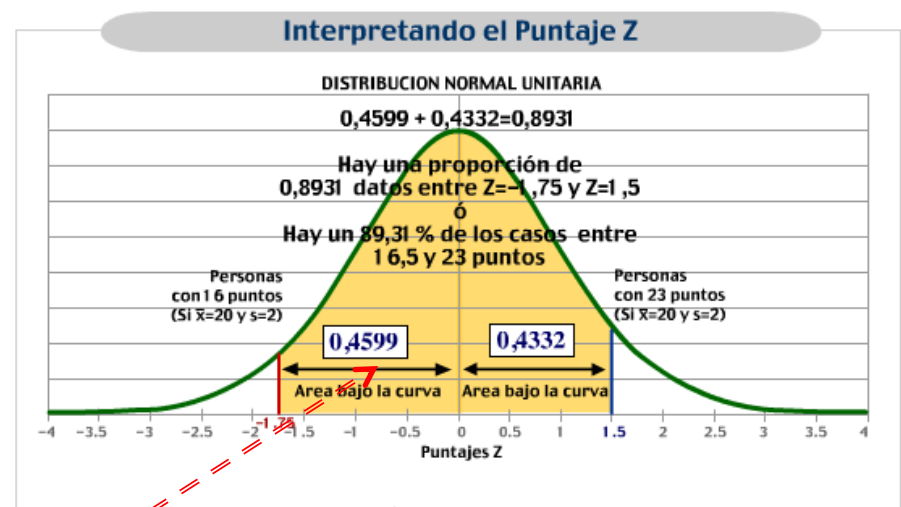


**Tabla de áreas bajo la
curva normal unitaria**

Interpretando el Puntaje Z



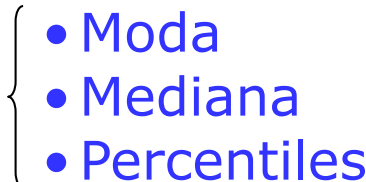
TABLAS DE PUNTAJES Z										
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4574	0.4584	0.4594	0.4604	0.4614	0.4625	0.4633	0.4643
1.8	0.4651	0.4659	0.4666	0.4674	0.4681	0.4688	0.4695	0.4702	0.4709	0.4716
1.9	0.4724	0.4731	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767	0.4772	0.4778
2.0	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817	0.4821	0.4826
2.1	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857	0.4861	0.4864



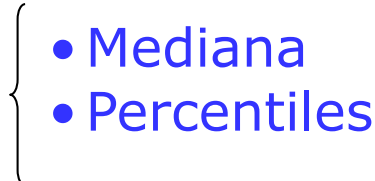
Haga clic en http://www.uc.cl/sw_educ/micssweb/html/pres5.htm
Para ver representación visual y profundizar

Elección de medidas de tendencia central y de dispersión

Variable **Nominal**  **Moda**

Variable **Ordinal**  

Variable **Continúa**:

- Con distribución desconocida o asimétrica  

- Con distribución simétrica y unimodal (Ej: Normal)
 

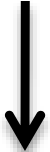
Verificación de hipótesis

Se utilizan una serie de procedimientos estadísticos, dósimas de comparación, dósimas de relación, etc.

Comparación

Entre:	Nominal - Ordinal	Intervalo - Razón
2 Grupos	N<40 Test Exacto de Fisher	Wilcoxon
	N>40 Chi cuadrado	Student Pareado Independiente
3 o más	Chi cuadrado	ANOVA Independiente Muestras repetidas

Análisis de relaciones

	Nominal	Ordinal	Intervalar-Razón
Nominal	Fi de contingencia		
Ordinal		rs Spearman	
Intervalar-Razón			r Pearson