

Estadística Descriptiva

Distribución y función de distribución de las
variables

Rodrigo Asun Inostroza

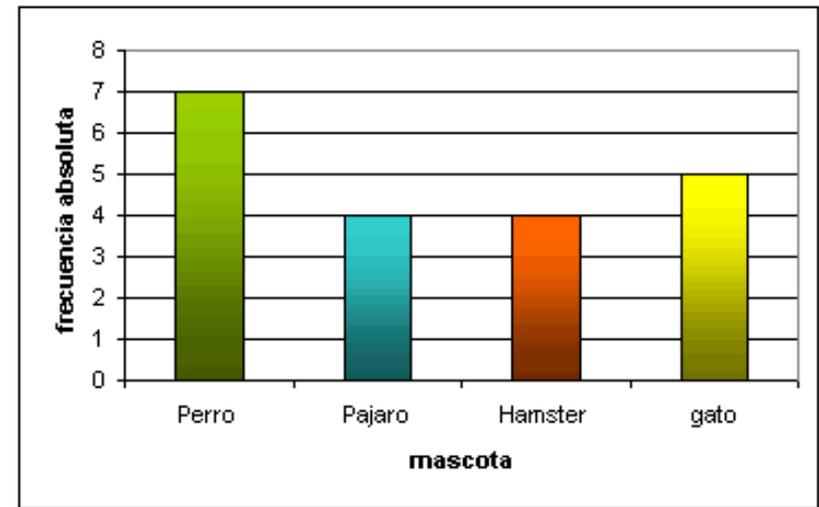
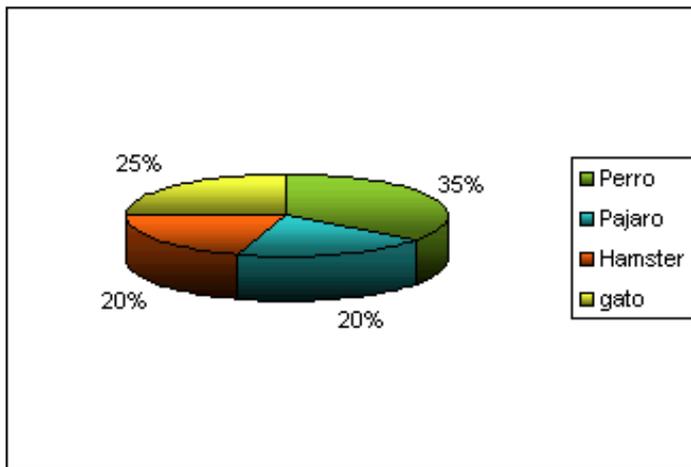
Temas a tratar

- Noción de “**Distribución de las variables**” aplicada a variables nominales, ordinales y de intervalo:
 - Distribuciones discretas y continuas.
- Conceptos de **Asimetría y Curtosis** de las variables.
- Noción de “**Función de Distribución**”:
 - Distribuciones empíricas versus teóricas.
- Función de **Distribución Normal Estándar** y sus propiedades:
 - Probabilidades bajo la curva normal, estandarización y puntajes Z.
 - Distribución de muestreo.

Noción de “Distribución de las Variables”

- Ordenamiento de las categorías de respuesta de una variable y su frecuencia de ocurrencia.
- Para una variable nominal se resume en su **tabla de frecuencia** y en un **gráfico** (usualmente de **torta** o barras)

Mascota	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada
Perro	7	.35	35 %
Pajaro	4	.20	20 %
Hamster	4	.20	20 %
gato	5	.25	25 %

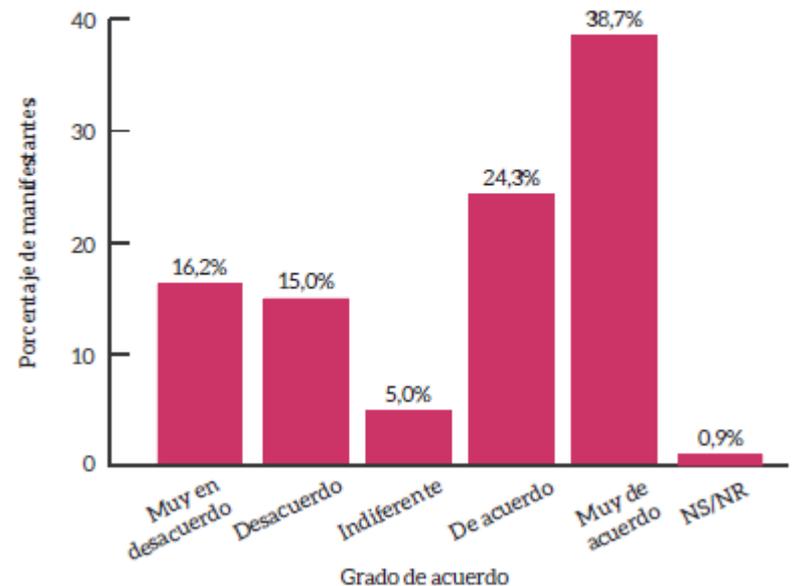


Noción de “Distribución de las Variables”

- Para una variable Ordinal se resume en su **tabla de frecuencia y en un gráfico** (usualmente barras, puesto que implica tácitamente orden de las variables).

	Frecuencia	Porcentaje	Porcentaje acumulado
Muy en desacuerdo	144	16.2	16.2
En desacuerdo	133	15.0	31.2
Indiferente	44	5.0	36.2
De acuerdo	215	24.3	60.5
Muy de acuerdo	342	38.7	99.1
NS / NR	8	0.9	100.0
Total	886	100	

Estoy participando en esta manifestación para defender mis intereses

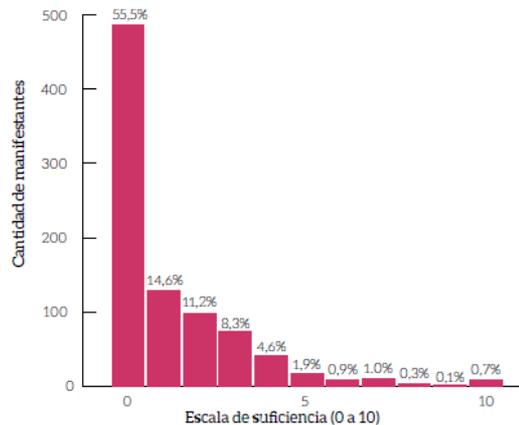


Noción de “Distribución de las Variables”

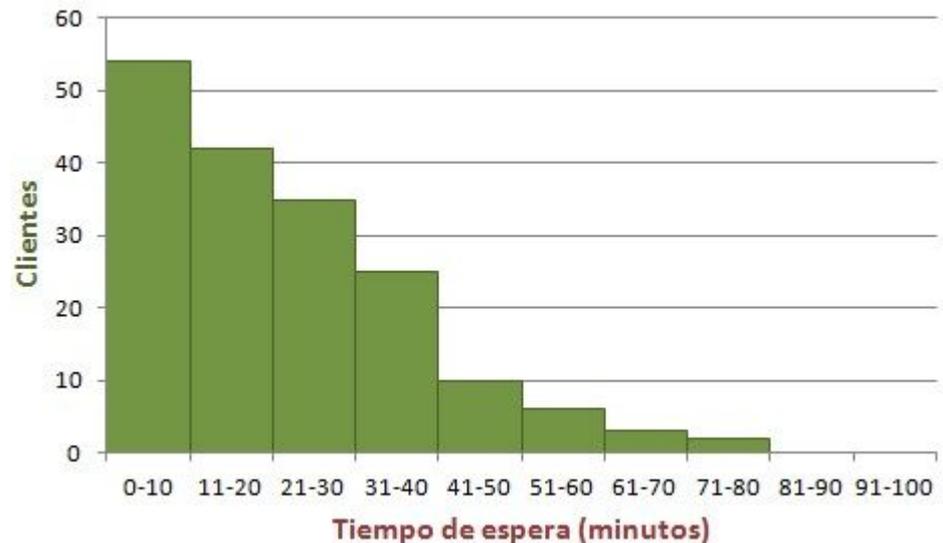
- Para una variable de Intervalo (Cuantitativa) se resume en su **tabla de frecuencia y en un gráfico** (usualmente Barras o Histograma, esto último cuando hay muchas categorías).

Gráfico de barras tradicional

El gobierno ha respondido a las movilizaciones con una nueva Agenda Social que incluye medidas en los siguientes temas: pensiones, salud y medicamentos, sueldo mínimo, tarifas eléctricas, entre otras. ¿En qué grado Ud. cree que son suficientes estas medidas? (considere 0 como insuficiente y 10 como suficiente).

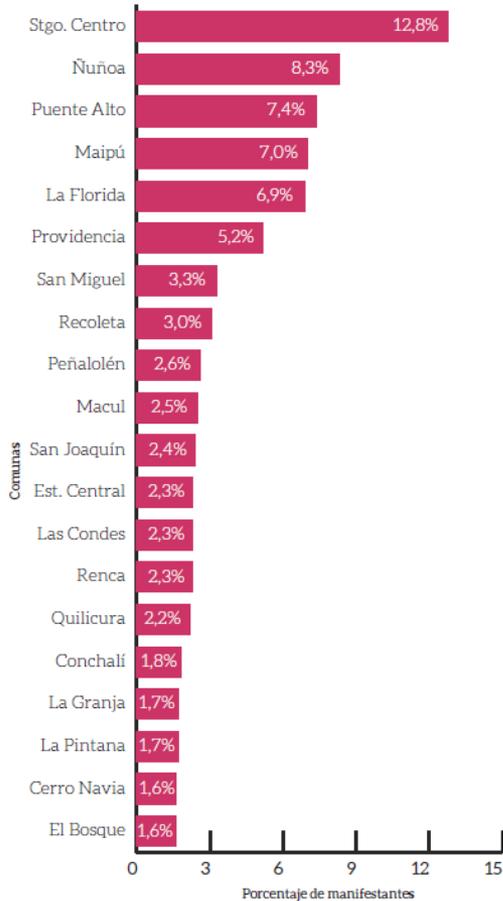


Histograma: agrupar variables continuas o discretas con muchas categorías



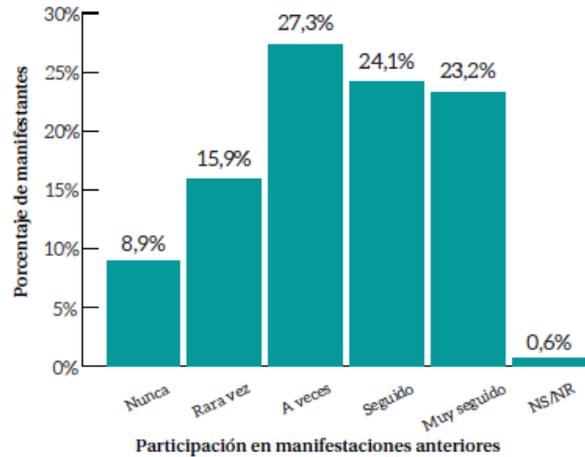
Distribuciones continuas y discretas

Top 20 Comunas



Variable nominal, **discreta**, con muchas categorías

Antes de este estallido social, en los últimos diez años, ¿con qué frecuencia usted había participado en una marcha o manifestación?



Variable Ordinal, **discreta**, con pocas categorías

Variable Intervalar, **continua**, con muchas categorías

¿Qué edad tiene?

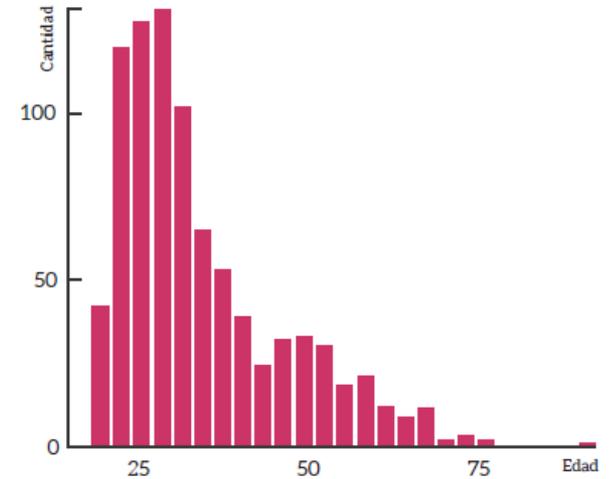
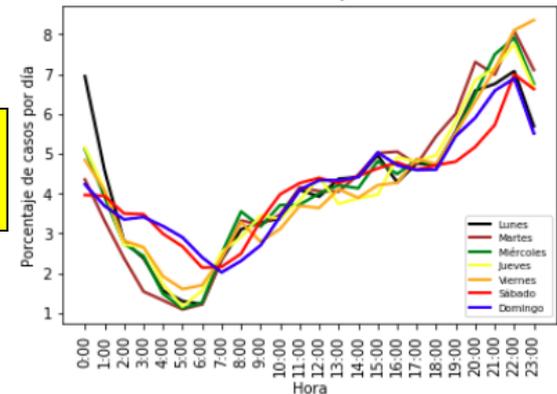


Figura 1: Histograma de edad.

Variable Intervalar, **discreta**, con muchas categorías

Figura 2. Distribución horaria, por día de la semana, de los casos de VIF física contra la mujer



Asimetría y Curtosis

- **Asimetría**: Medida en que la distribución de una variable se **aleja de una distribución simétrica** a ambos lados de su punto central.
- **Curtosis**: Medida del **grado de agrupamiento** de los datos de una distribución.
 - Solo tienen pleno sentido con variables de intervalo o superiores, pero se aplica a variables ordinales con muchas categorías.

Coeficiente de
Asimetría de Fisher:

$$CA_F = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N \cdot S_x^3}$$

Coeficiente de
Curtosis:

$$g_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4 \cdot n_i}{N S_x^4}$$

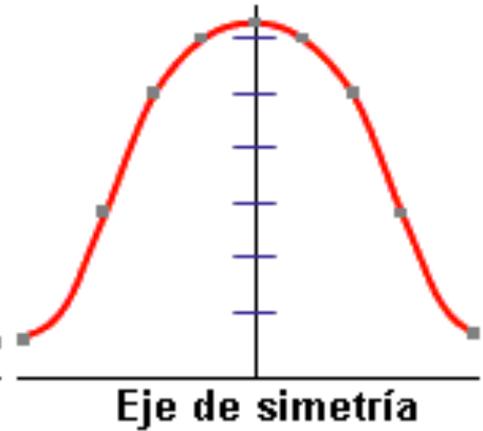
n_i la frecuencia absoluta de x_i

Asimetrías Moderadas

Curva de asimetría Negativa



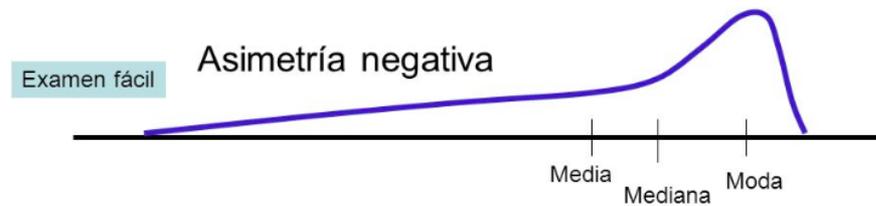
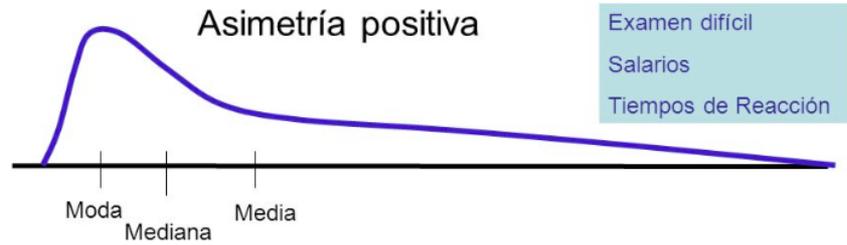
Curva simétrica

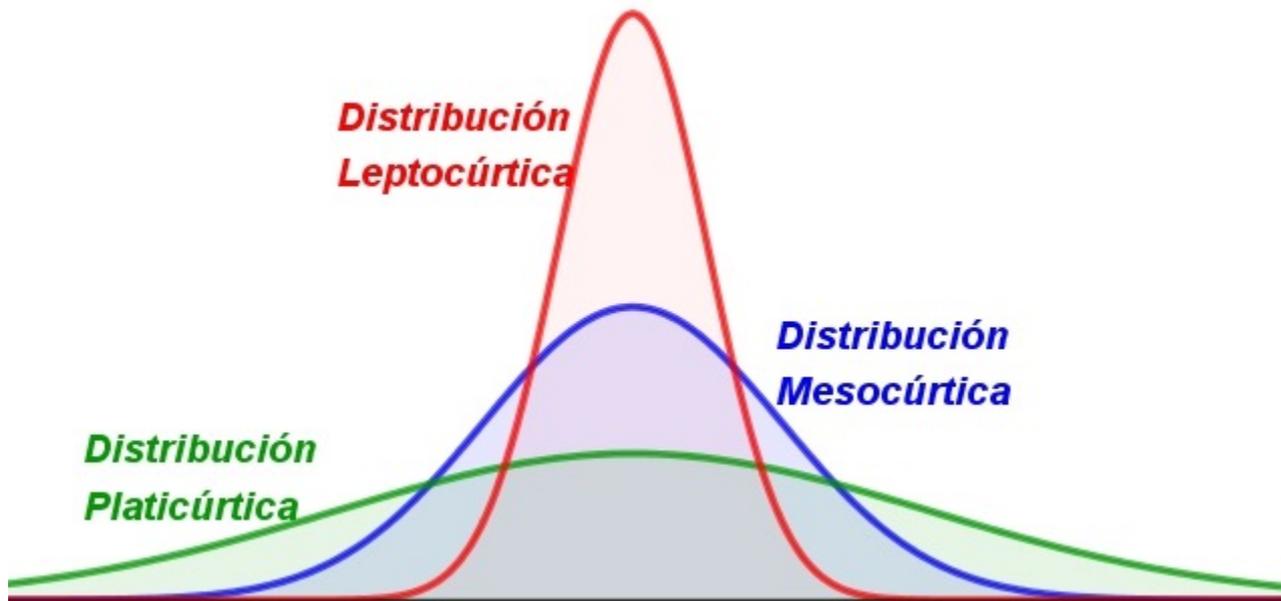


Curva de asimetría Positiva



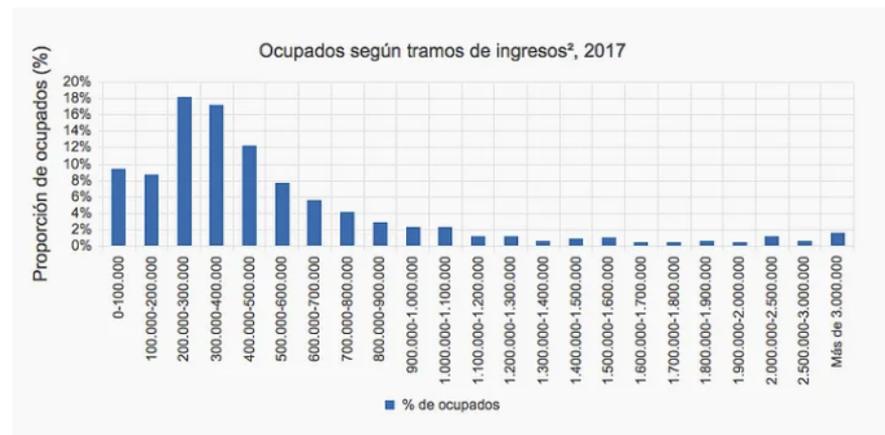
Asimetrías Extremas





Distribución de Ingresos en Chile

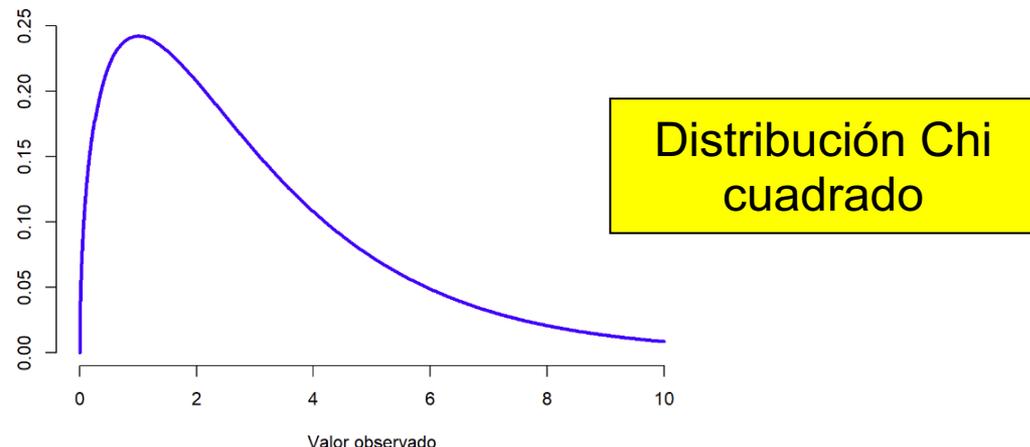
¿Cómo describiría la distribución de los ingresos en Chile?

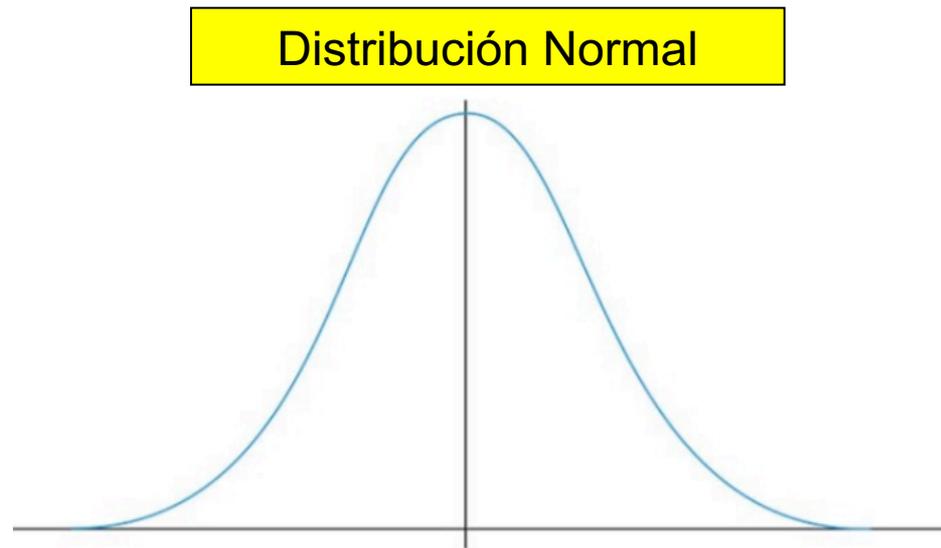
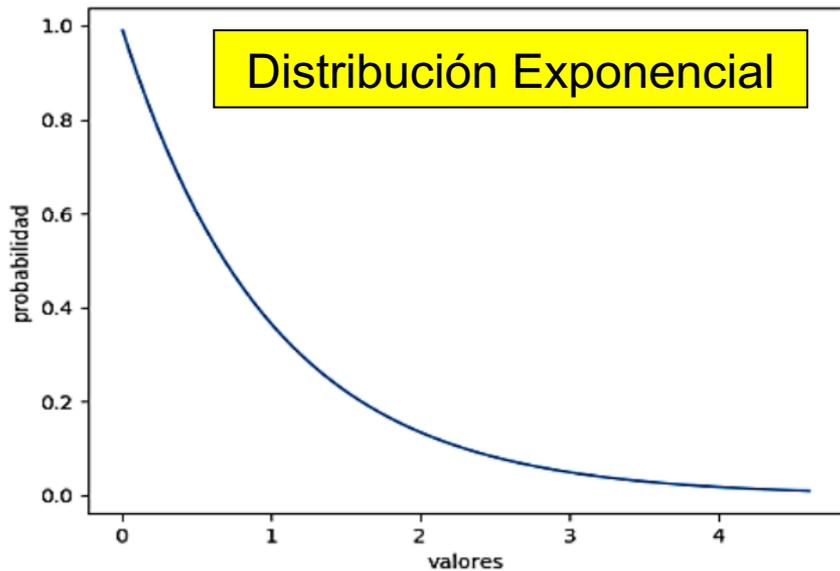


Distribuciones empíricas y teóricas

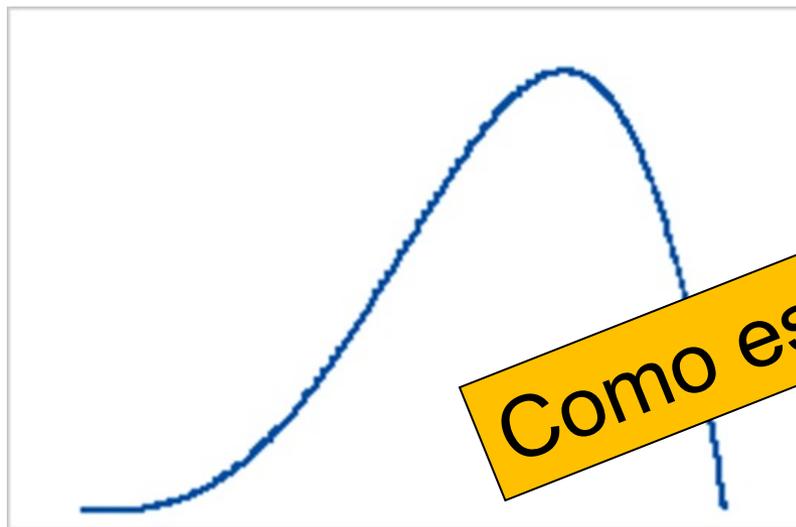
- Las distribuciones **empíricas existen...** son la distribución de una variable real.
- Las distribuciones teóricas son **formulaciones ideales matemáticas**. No existen en la vida real.
- Expresan la distribución de un conjunto de números en términos de su probabilidad de ocurrencia. Son **Distribuciones de Probabilidad**.

- Ejemplos:

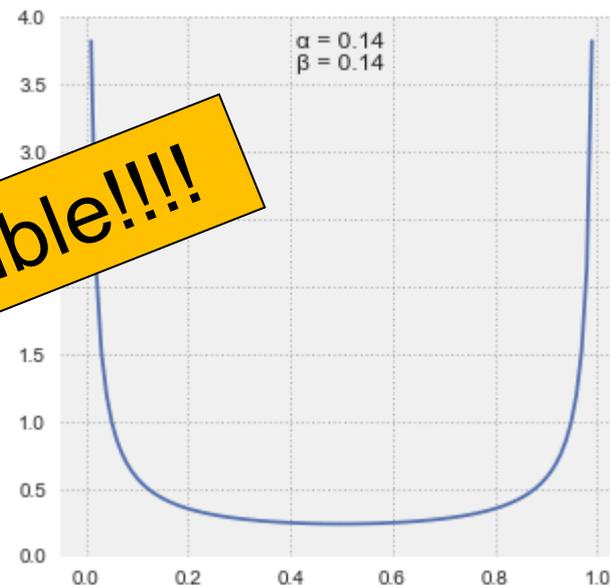




Distribución Beta



Distribución Beta



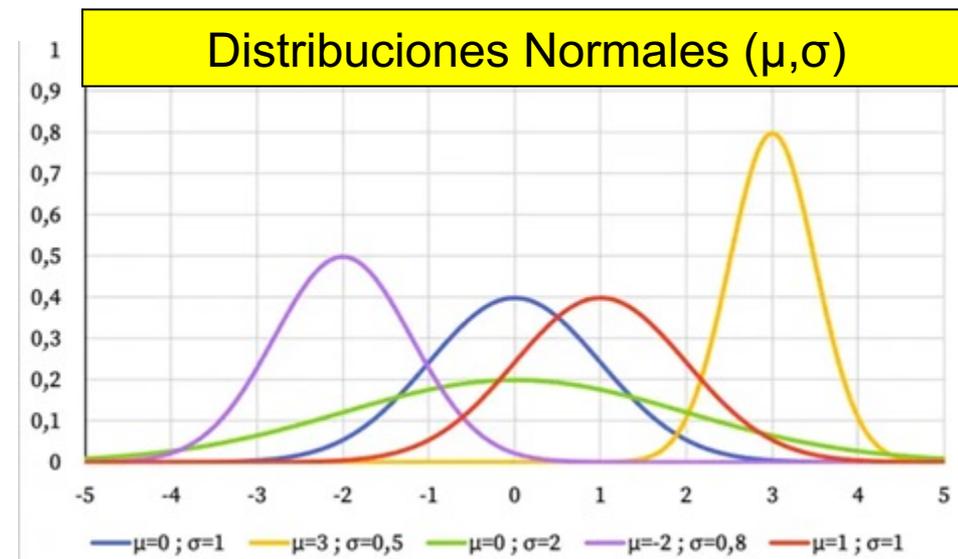
Como es posible!!!!

Familias de distribuciones teóricas

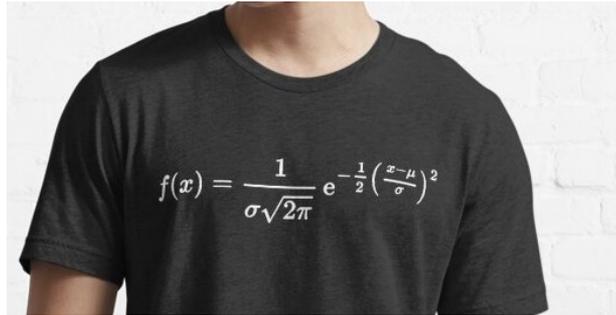
- Las distribuciones teóricas **rara vez son únicas**, en realidad son:
 - Familias de Distribuciones**, que se producen como resultado de una **Función Matemática** llamada “Función de Densidad”.
 - Que puede asumir **distintos parámetros**.
- Ejemplo:**

Función de densidad de la distribución normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

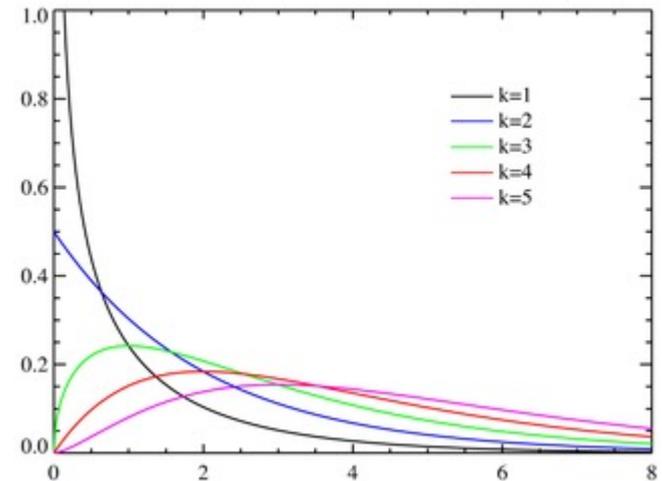


Ejemplos de funcionamiento de funciones de densidad



Función de Distribución Chi cuadrado

$$\frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$



<https://www.medcalc.org/manual/chidist-function.php>

https://es.wikipedia.org/wiki/Archivo:PDF_Beta_Distribution_Animation.gif

¿Para que sirven las distribuciones teóricas?

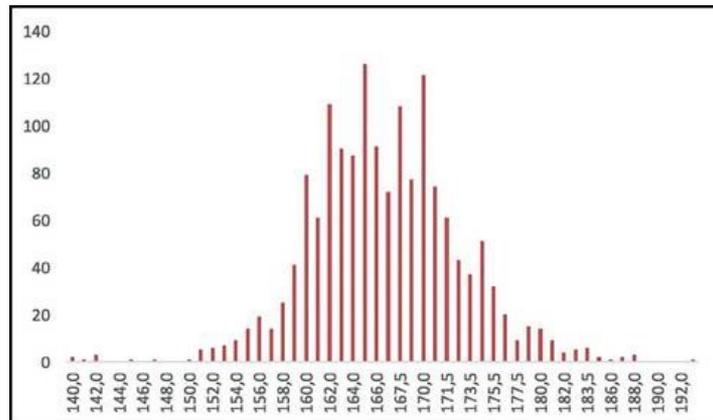
- Cuando una distribución empírica se parece a una distribución teórica...
 - Podemos utilizar las **propiedades de la distribución teórica** para **saber cosas de la distribución empírica**.
 - Como fue causada la distribución empírica.
 - Probabilidad de cierto resultado.
 - Hay ciertas situaciones en que **necesariamente la realidad** asume una determinada distribución teórica.
 - Ejemplo: **distribución de muestreo**.

Ejemplo: Propiedades de la curva normal.

- ¿**Cómo se produce** una curva normal?
 - Cuando un determinado resultado es producto de **muchas causas pequeñas**, ninguna de las cuales es mucho más importante que las otras.

<https://youtube.com/shorts/TwctT3Ncm1w?feature=share>

- Ejemplo: Estatura una población.

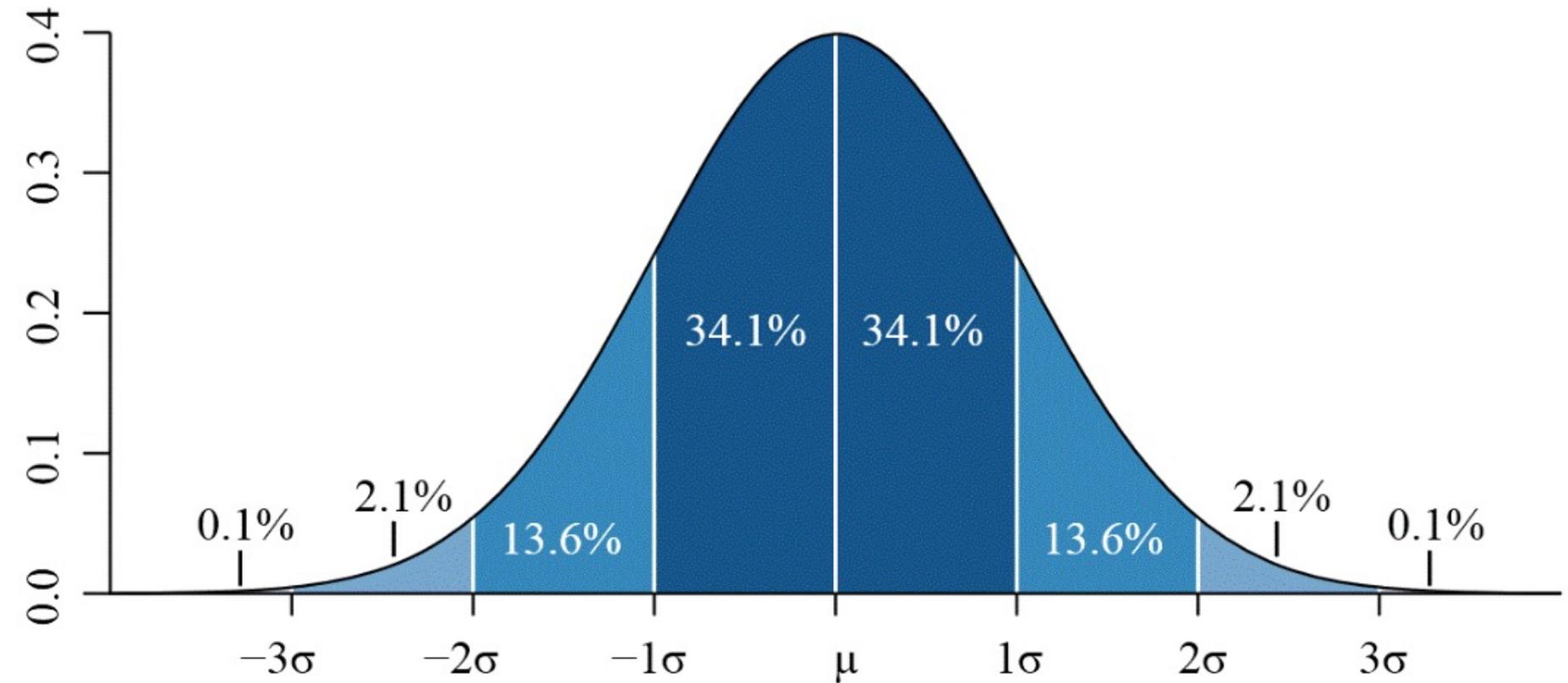


Estaturas Chile Siglo XIX

Probabilidades bajo la curva Normal

- Si una curva empírica se parece a una Normal, se **pueden utilizar las probabilidades conocidas de distribución de la Normal** para conocer las probabilidades empíricas.
 - Ejemplo: la distribución de las notas en un curso se parece a una Normal (4.5, 1.2)...
 - (Es decir, con media 4,5 y desviación estándar 1,2)
 - ¿Qué porcentaje de los estudiantes sacó nota superior a 4,5?
 - ¿Qué porcentaje de los estudiantes sacó nota inferior a 3,3?
 - ¿Qué porcentaje de los estudiantes sacó nota 6,9 o más?

Recuerde: Probabilidades bajo la curva Normal

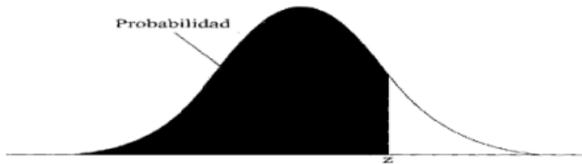


Generalizando el Cálculo de probabilidades bajo la curva normal

- Puntuación Z: **N(0,1)**
- Transformar a puntuación Z:
 - Nota de Juanito: 5,6
 - Distribución N(4.5,1.2)
- Fórmula:
$$z = \frac{x - \bar{x}}{s}$$
- Puntuación Z Juanito: 0,92.
- ¿**A que probabilidad corresponde?**

Cálculo de probabilidades bajo la curva normal

- Tabla Z



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861

- Web: <https://www.calculator.net/z-score-calculator.html>

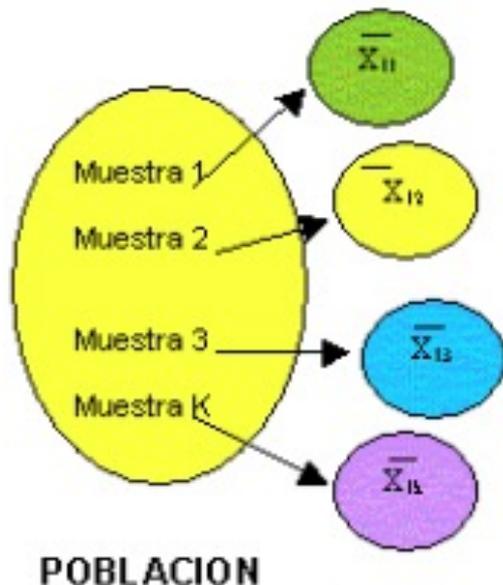
Para finalizar... **distribución de muestreo** de los promedios

- Cuando uno extrae infinitas muestras de tamaño grande (>100) de un universo cualquiera.
- Y en cada muestra calcula el promedio obtenido.
- La distribución de esos promedios tendrá Distribución Normal, con una media igual al promedio del universo y una desviación estándar igual a la desviación estándar del universo, dividida por la raíz cuadrada del tamaño de las muestras.
- Es decir, tendrá **distribución $N(\mu, \sqrt{n})$** .

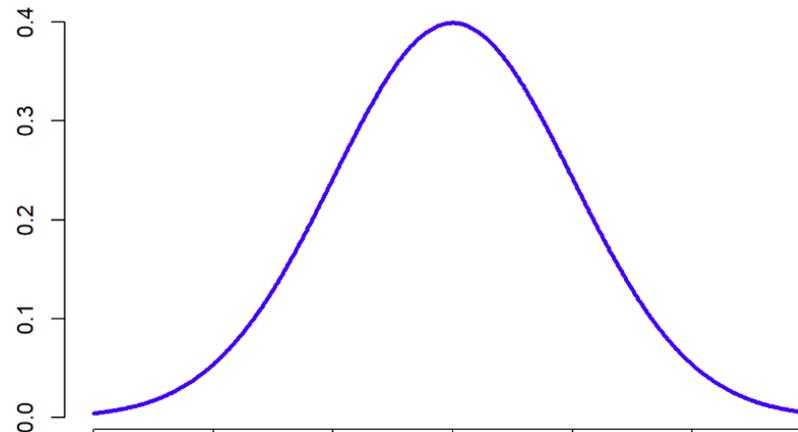
Ejemplo y Representación gráfica

- Del Universo de Chilenos y chilenas, **extraigo infinitas muestras** de 150 personas cada una y en cada una de ellas mido el nivel de Machismo.
- En cada muestra calculo el promedio de Machismo de la muestra.
- Si calculo el promedio de promedios de Machismo, realizo un gráfico de su distribución y calculo la desviación estándar de los promedios...

Problema!!!!



Obtendré una distribución $N(\mu, \sigma/\sqrt{n})$



Implica que podré conocer: el nivel de machismo de la población y su desviación estándar!!!

FIN