Estadística General

Tema 4: Distribución de Probabilidades

Distribución Normal, Teorema Central del límite.

Prof. José G. Páez

Distribución normal o de Gauss

- Aparece de manera natural:
  - Errores de medida.
  - Distancia de frenado.
  - Altura, peso, propensión al crimen…
  - Distribuciones binomiales con n grande (n>30) y ‘p ni pequeño’ (np>5) ‘ni grande’ (nq>5).

- Está caracterizada por dos parámetros: La media, μ, y la desviación típica, σ.

- Su función de densidad es:

Función de Densidad

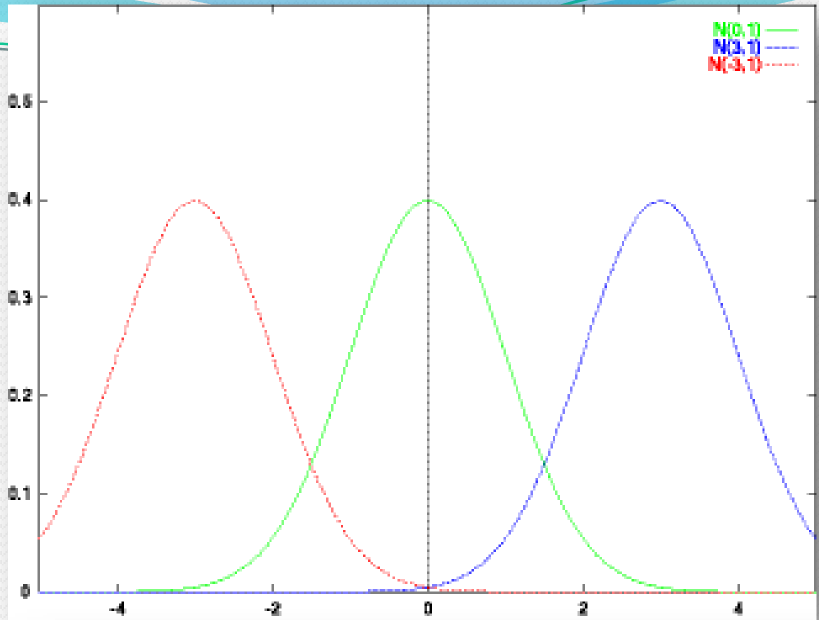
$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

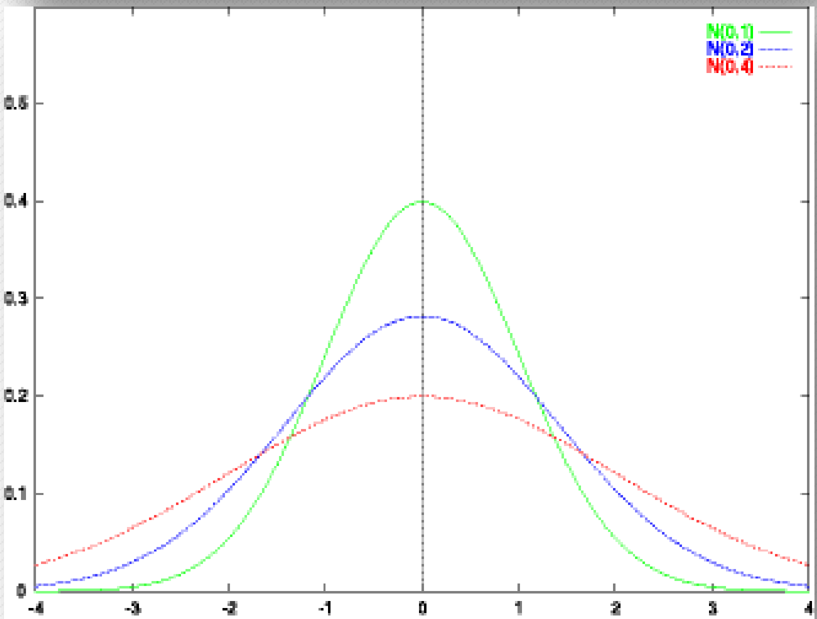
μ	media	π	= 3,1415...
σ	desv. típica	e	= 2,7182...
σ^2	varianza	x	abscisa

N(μ, σ): Interpretación geométrica

- Puedes interpretar la media como un factor de traslación.

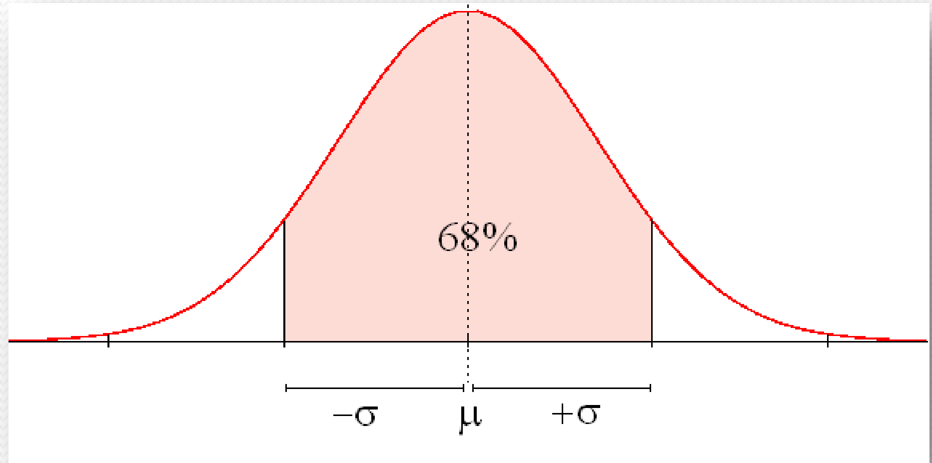
- Y la desviación típica como un factor de escala, grado de dispersión,…



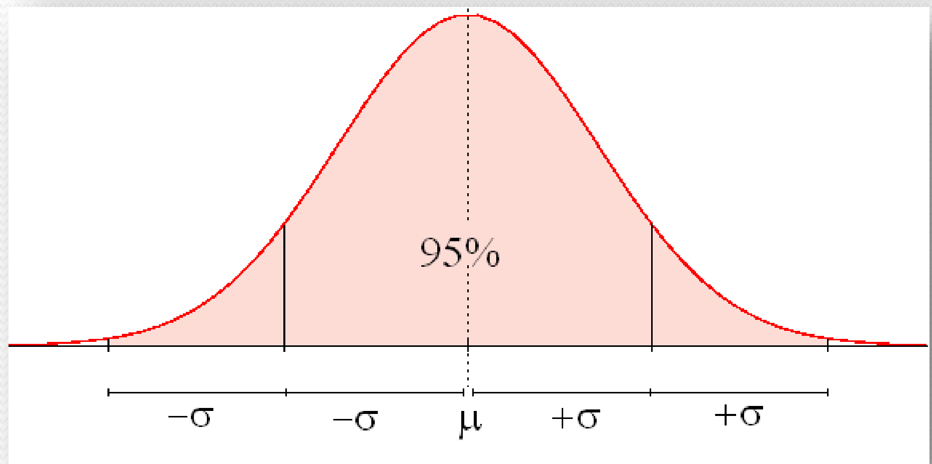


$N(\mu, \sigma)$: Interpretación probabilista

- Entre la media y una desviación típica tenemos siempre la misma probabilidad: aprox. 68%



- Entre la media y dos desviaciones típicas aprox. 95%



Algunas características

- La función de densidad es simétrica, mesocúrtica y unimodal.
  - Media, mediana y moda coinciden.

- Los puntos de inflexión de la fun. de densidad están a distancia **σ** de μ.

- Si tomamos intervalos centrados en μ, y cuyos extremos están…
  - a distancia **σ**, ➔ tenemos probabilidad **68%**
  - a distancia **2 σ**, ➔ tenemos probabilidad **95%**
  - a distancia **2'5 σ** ➔ tenemos probabilidad **99%**

- No es posible calcular la probabilidad de un intervalo simplemente usando la primitiva de la función de densidad, ya que no tiene primitiva expresable en términos de funciones 'comunes'.

- Todas las distribuciones normales N(μ, σ), pueden ponerse mediante una traslación μ, y un cambio de escala σ, como **N(0,1)**. Esta distribución especial se llama **normal tipificada**.
  - Justifica la técnica de tipificación, cuando intentamos comparar individuos diferentes obtenidos de sendas poblaciones normales.

Tipificación

- Dada una variable de media μ y desviación típica σ, se denomina valor tipificado,z, de una observación x, a la distancia (con signo) con respecto a la media, medido en desviaciones típicas, es decir

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

- En el caso de variable X normal, la interpretación es clara: Asigna a todo valor de $N(\mu, \sigma)$, un valor de $N(0,1)$ que deja **exactamente** la misma probabilidad por debajo.

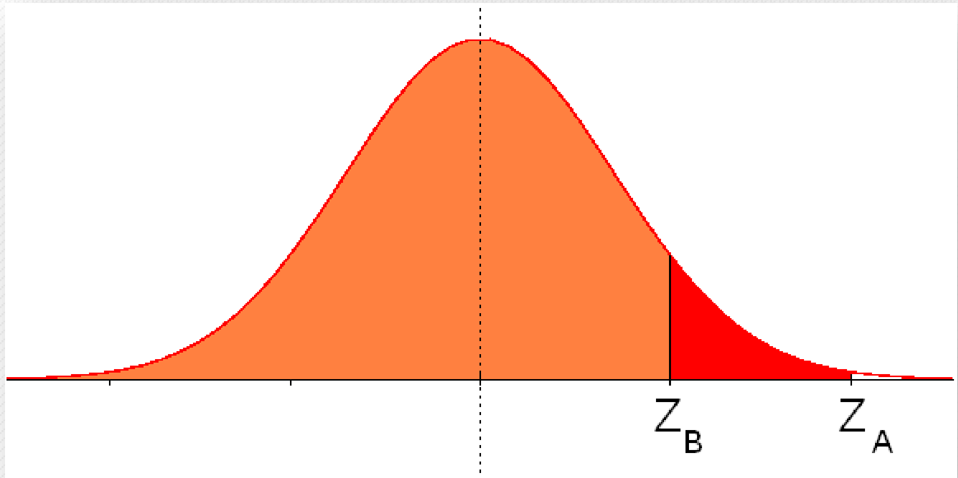
- Nos permite así comparar entre dos valores de dos distribuciones normales diferentes, para saber cuál de los dos es más extremo.

Ejemplo

- Se quiere dar una beca a uno de dos estudiantes de sistemas educativos diferentes. Se asignará al que tenga mejor expediente académico.
  - El estudiante A tiene una calificación de 8 en un sistema donde la calificación de los alumnos se comporta como N(6,1).
  - El estudiante B tiene una calificación de 80 en un sistema donde la calificación de los alumnos se comporta como N(70,10).
- **Solución**
  - No podemos comparar directamente 8 puntos de A frente a los 80 de B, pero como ambas poblaciones se comportan de modo normal, podemos tipificar y observar las puntuaciones sobre una distribución de referencia N(0,1)

$$z_A = \frac{x_A - \mu_A}{\sigma_A} = \frac{8-6}{1} = 2$$

$$z_B = \frac{x_B - \mu_B}{\sigma B} = \frac{80-70}{10} = 1$$



  - Como $Z_A > Z_B$, podemos decir que el porcentaje de compañeros del mismo sistema de estudios que ha superado en calificación el estudiante A es mayor que el que ha superado B. Podríamos pensar en principio que A es mejor candidato para la beca.

¿Por qué es importante la distribución normal?

- Las propiedades que tiene la distribución normal son interesantes, pero todavía no hemos hablado de por qué es una distribución especialmente importante.

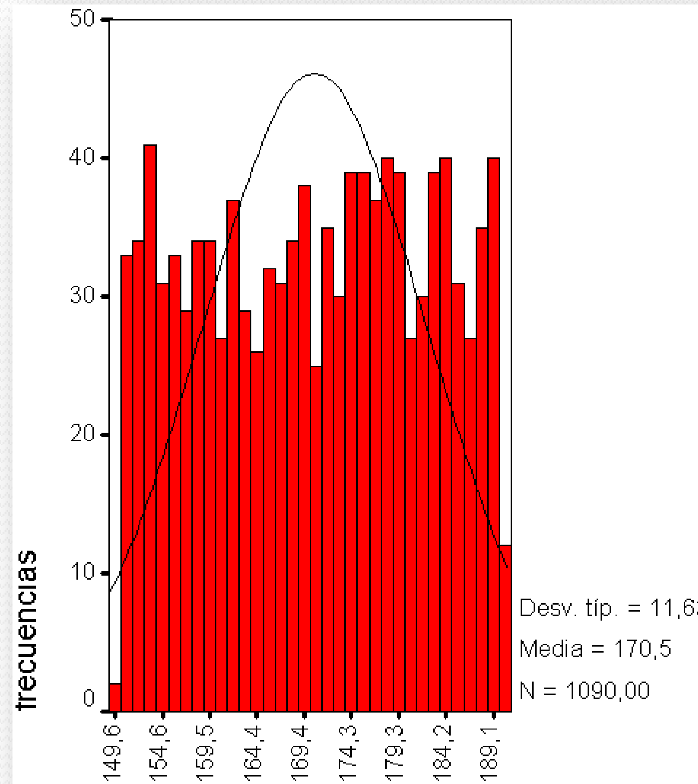
- La razón es que aunque una variable aleatoria no posea distribución normal, ciertos estadísticos/estimadores calculados sobre muestras elegidas al azar sí que poseen una distribución normal.

- Es decir, tengan las distribución que tengan nuestros datos, los 'objetos' que resumen la información de una muestra, posiblemente tengan distribución normal (o asociada).

Veamos aparecer la distribución normal

- Como ilustración mostramos una variable que presenta valores distribuidos más o menos uniformemente sobre el intervalo 150-190.

- Como es de esperar la media es cercana a 170. El histograma no se parece en nada a una distribución normal con la misma media y desviación típica.



- A continuación elegimos aleatoriamente grupos de 10 observaciones de las anteriores y calculamos el promedio.

- Para cada grupo de 10 obtenemos entonces una nueva medición, que vamos a llamar promedio muestral.

- Observa que las nuevas cantidades están más o menos cerca de la media de la variable original.

- Repitamos el proceso un número elevado de veces. En la siguiente página estudiamos la distribución de la nueva variable.

Muestra		
1ª	2ª	3ª
185	190	179
174	169	163
167	170	167
160	159	152
172	179	178
183	175	183
188	159	155
178	152	165
152	185	185
175	152	152

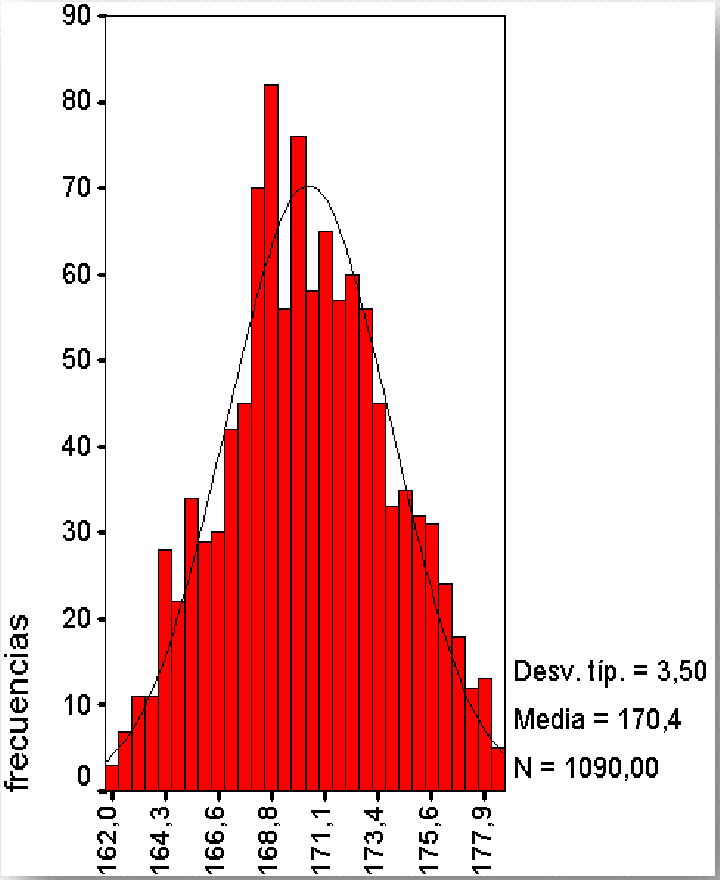
173 169 168 ...

- La distribución de los promedios muestrales sí que tiene distribución aproximadamente normal.

- La media de esta nueva variable (promedio muestral) es muy parecida a la de la variable original.

- Las observaciones de la nueva variable están menos dispersas. Observa el rango. Pero no sólo eso. La desviación típica es aproximadamente 'raiz de 10' veces más pequeña. Llamamos error estándar a la desviación típica de esta nueva variable.

- **Nada** de lo anterior **es casualidad**.



Teorema central del límite

- Dada una v.a. **cualquiera**, si extraemos muestras de tamaño n, y calculamos los promedios muestrales, entonces:
- dichos promedios tienen distribución aproximadamente normal;
- La media de los promedios muestrales es la misma que la de la variable original.
- La desviación típica de los promedios disminuye en un factor "raíz de n" (error estándar).
- Las aproximaciones anteriores se hacen exactas cuando n tiende a infinito.
  - Este teorema justifica la importancia de la distribución normal.
  - Sea lo que sea lo que midamos, cuando se promedie sobre una muestra grande (**n>30**) nos va a aparecer de manera natural la distribución normal.