

**SEGUNDO ENCUENTRO TÉCNICO  
SOBRE LA ESTRUCTURACIÓN DE PROYECTOS DE ASOCIACIÓN  
PÚBLICO-PRIVADA**

# **Elementos Conceptuales para la Cuantificación del Riesgo**

**Sergio Hinojosa**

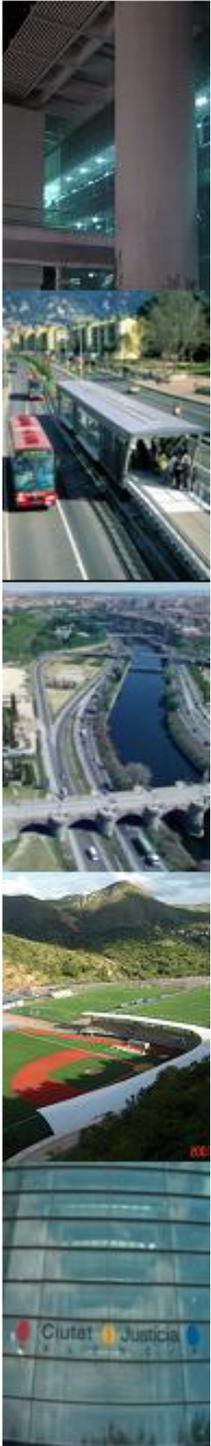
PIAPPEM

**28,29 y 30 de Abril de 2009**

**Guanajuato, Ciudad de Guanajuato**

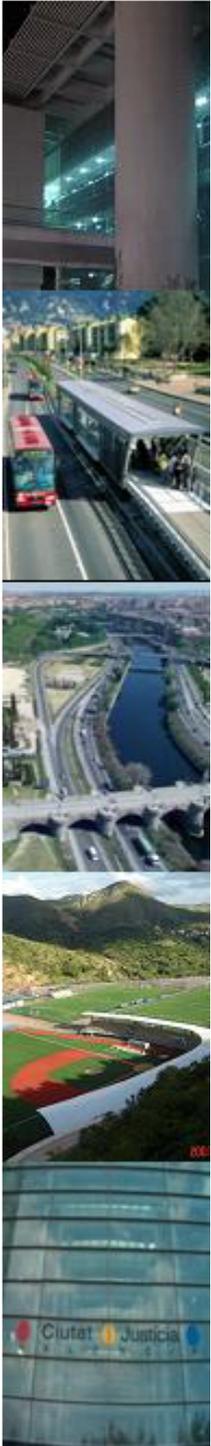
# Índice de Temas

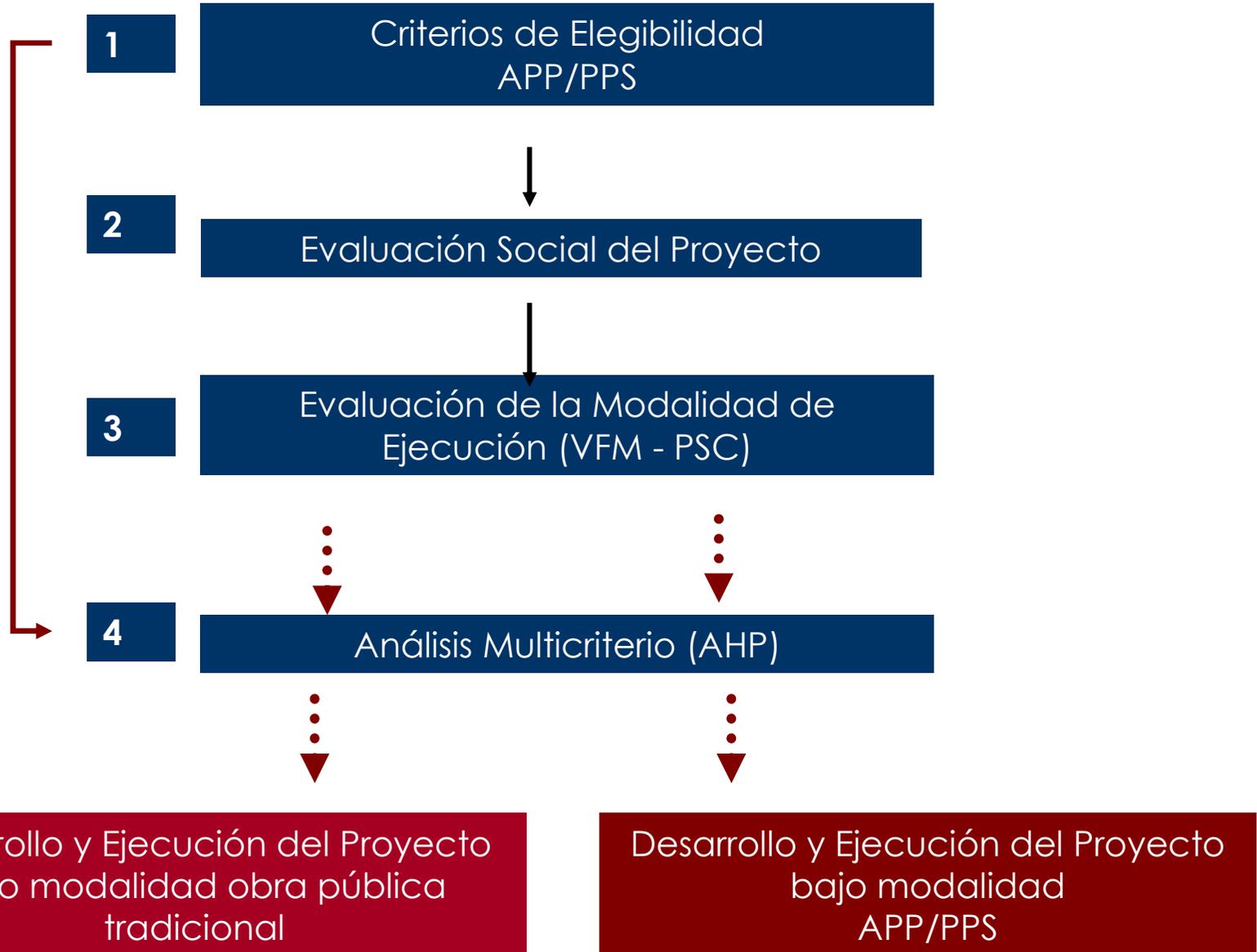
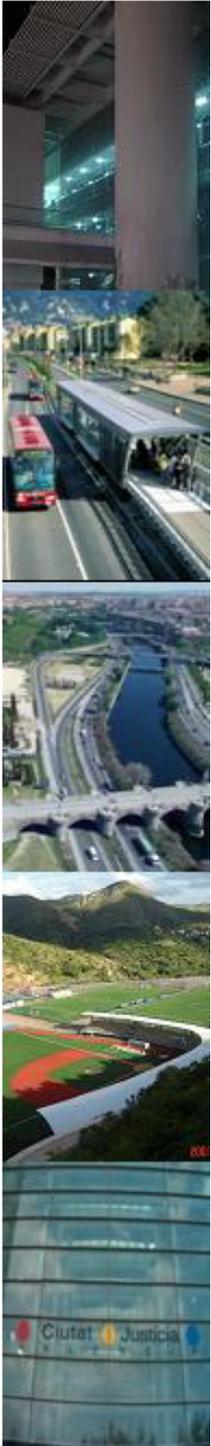
1. Contexto del CPP Estados Mexicanos
2. Consistencia de flujos y el equivalente cierto
3. Valor esperado, varianza, Histograma y definición de riesgo
4. Supuestos del enfoque riesgo – retorno o media varianza
5. Actitudes frente al riesgo
6. La distribución Normal
7. Pruebas de Normalidad
8. Definición de Percentil
9. Principios de la simulación de Montecarlo
10. Ejercicio aplicado de valoración de riesgos CB
11. La simulación Bootstrap



# Enfoque

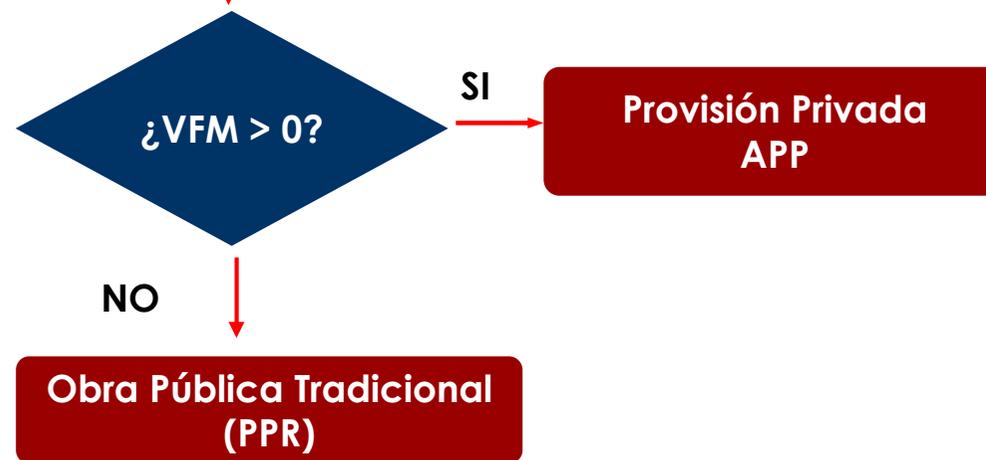
1. Generar Lenguaje
2. Generar Intuición
3. Destacar algunas palabras claves
4. Destacar algunos conceptos claves
5. Facilitar la comprensión del Diploma APP/PIAPPEM



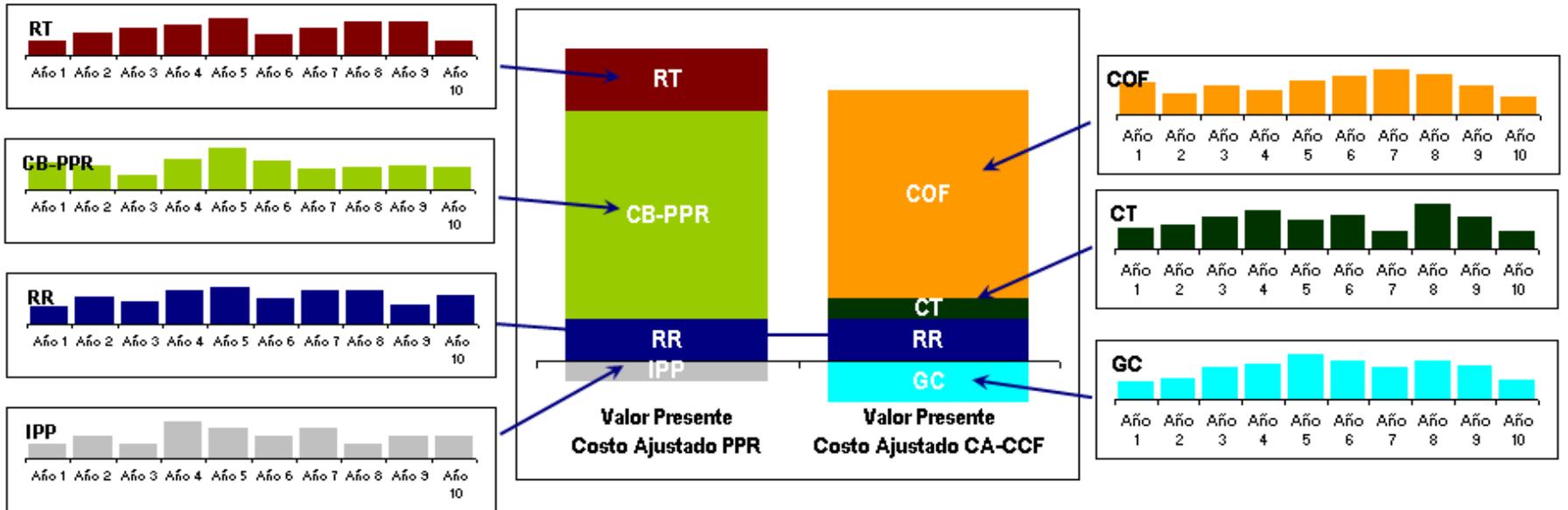


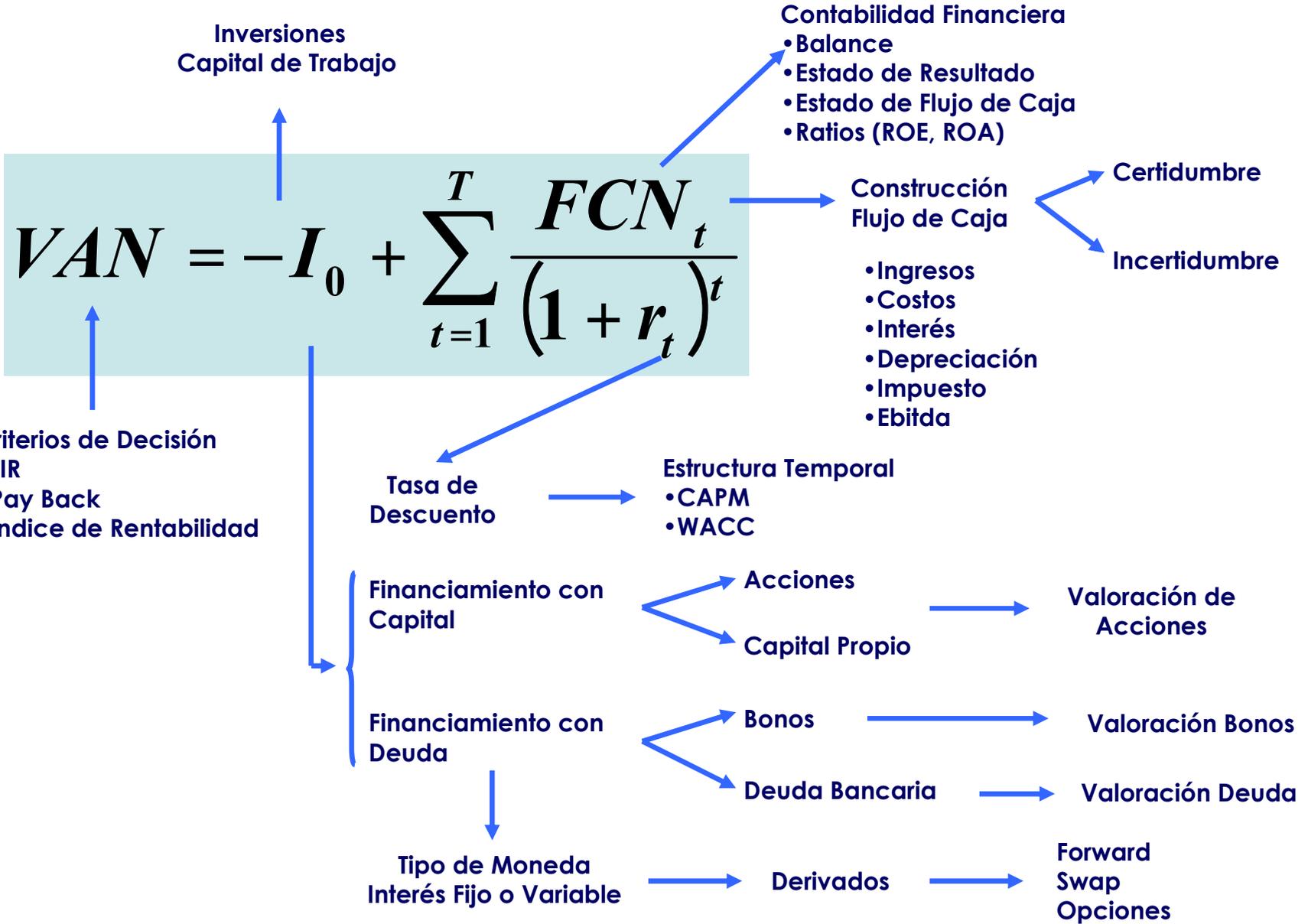
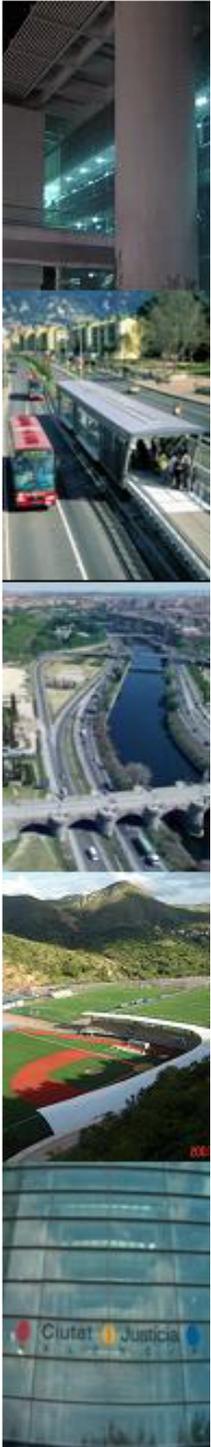
# Valor por Dinero (VFM)

$$\text{VFM} = \text{Costo de la provisión pública ajustado por riesgo} - \text{Costo de la provisión privada ajustado por riesgo}$$



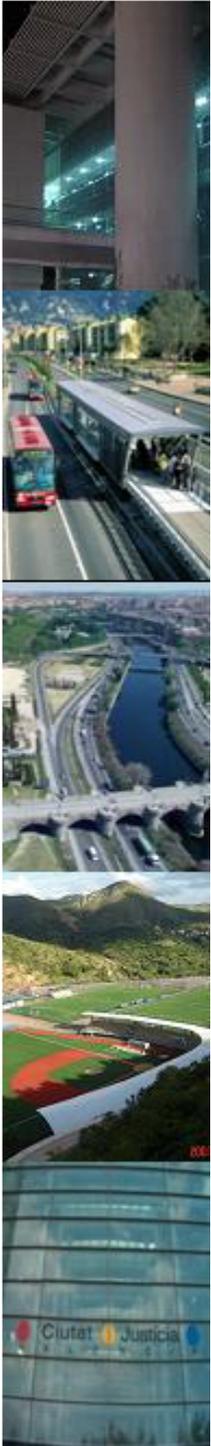
# En Valor Actual Neto

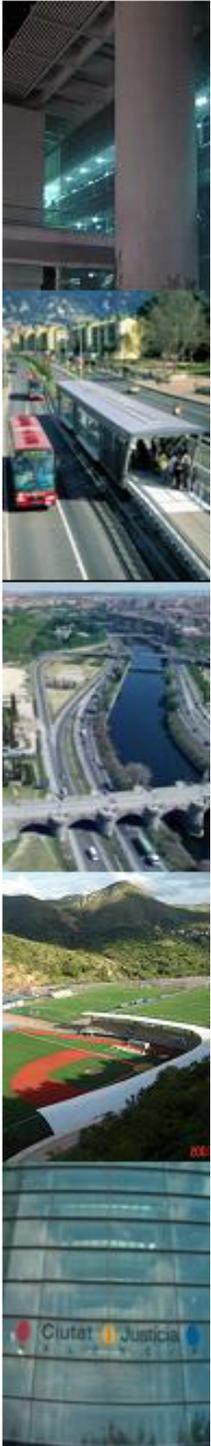




# Clave: Consistencia

1. Inflación: tasas nominales vs reales
2. Impuestos ( con / sin IVA)
3. Puntos de vista del Proyecto (WACC) , puntos de vista del accionistas o inversionista (CAPM)
4. Flujos sin riesgos tasa libre de riesgo, Flujos con riesgo tasa ajustada por riesgo



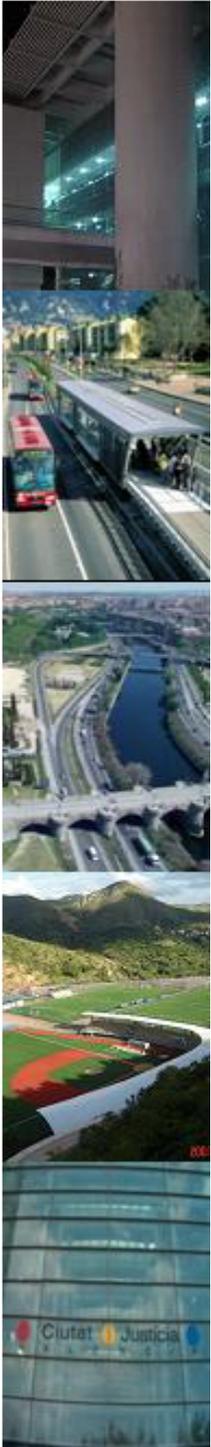


## Equivalente Cierto en Evaluación de Inversiones de Capital

La relación entre equivalente cierto (certainty-equivalent) y la tasa de descuento ajustada por riesgo fue discutida por primer vez por Robichek y Myers (1966) para el caso de un periodo. El caso para flujos de caja multiperiodo fue propuesto y desarrollado por Bogue y Roll (1974).

A continuación se presenta el caso para un periodo. Cuando calculamos el valor presente de un proyecto o activo, el riesgo se puede tomar en cuenta de dos maneras alternativas:

- i. Tasa
- ii. Flujo



## Equivalente Cierto en Evaluación de Inversiones de Capital

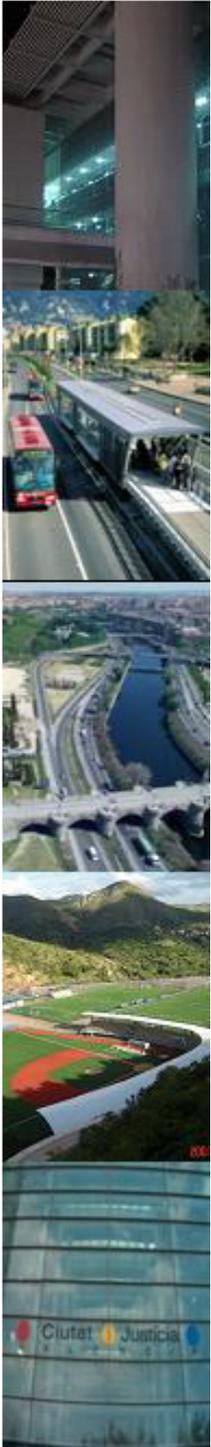
Es posible descontar el flujo de caja esperado por la tasa de descuento ajustada por riesgo:

$$VP = \frac{F_1}{(1 + r)}$$

O alternativamente se puede descontar el equivalente cierto del flujo de caja (FEC) por la tasa interés libre de riesgo.

$$VP = \frac{FEC_1}{(1 + r_f)}$$

El Flujo de caja equivalente cierto puede ser obtenido a través de modelo de activos de capital (CAPM) propuesto en los trabajos seminales de Sharpe (1964) y Lintner (1965).



## Equivalente Cierto en Evaluación de Inversiones de Capital

El modelo CAPM se representa como sigue:

$$r = r_f + \beta(r_m - r_f) \quad \text{lo que es equivalente a} \quad 1 + r = 1 + r_f + \beta(r_m - r_f)$$

En consecuencia, es posible expresar:

$$\frac{F_1}{VP} = 1 + r_f + \beta(r_m - r_f) \quad (1)$$

Es ampliamente conocido que el valor de  $\beta$  se calcula como la covarianza entre el retorno del activo y el retorno de mercado dividido por la varianza de mercado [Sharpe (1964), Lintner (1965)]:

$$\beta = \frac{\text{COV}(r, r_m)}{\sigma_m^2} = \frac{\text{COV}\left(\frac{F_1}{VP} - 1, r_m\right)}{\sigma_m^2}$$

## Equivalente Cierto en Evaluación de Inversiones de Capital

La cantidad  $F_1$  es el flujo de caja futuro y por lo tanto incierto. Pero  $VP$  es el valor presente del activo. Este valor no es desconocido y, por lo tanto no covaría con  $r_m$ . En consecuencia la expresión para  $\beta$  puede ser descrita cómo:

$$\beta = \frac{\text{cov}(F_1, r_m)}{VP \times \sigma_m^2}$$

Sustituyendo la expresión anterior en (1) se tiene:

$$\frac{F_1}{VP} = 1 + r_f + \frac{\text{cov}(F_1, r_m)}{VP} \times \frac{r_m - r_f}{\sigma_m^2}$$

# Equivalente Cierto en Evaluación de Inversiones de Capital

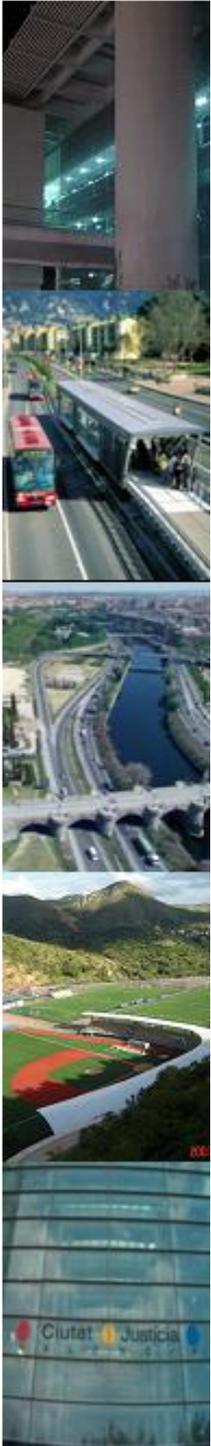
La expresión  $r_m - r_f / \sigma_m^2$  es el premio de riesgo esperado en el mercado por unidad de varianza. También es denominada el precio de mercado del riesgo [Sharpe (1964), Hull (2000)] y es escrita como  $\lambda$  (lambda).

Entonces:

$$\frac{F_1}{VP} = 1 + r_f + \frac{\lambda \text{cov}(F_1, r_m)}{VP}$$

Multiplicando ambos términos por VP y reordenando se tiene:

$$VP = \frac{F_1 - \lambda \text{cov}(F_1, r_m)}{1 + r_f}$$

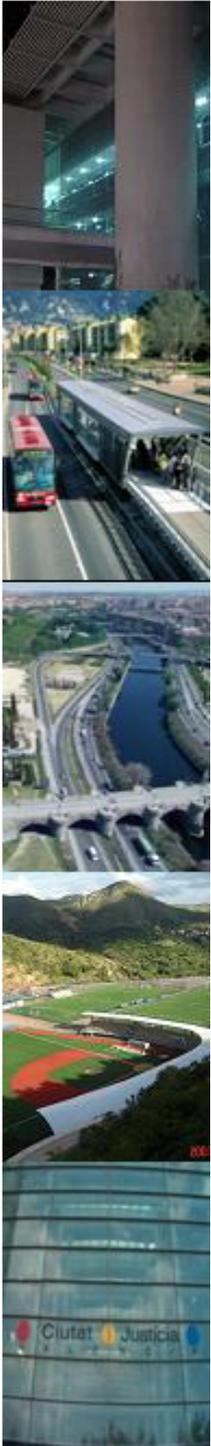


## Equivalente Cierto en Evaluación de Inversiones de Capital

La expresión anterior es el equivalente cierto del modelo de valoración de activos (CAPM). Nos indica que si el activo es libre de riesgo, la  $\text{cov}(F_1, r_m)$  es cero, y entonces el flujo  $F_1$  se descuenta simplemente a la tasa libre de riesgo. Pero, si el activo es riesgoso debemos descontar el equivalente cierto del flujo de caja  $F_1$ . Este descuento que debemos hacer de  $F_1$  depende del precio de mercado del riesgo y de la covarianza entre los flujos de caja del proyecto y el retorno de mercado.

Una aproximación a la expresión anterior es la valoración de la estimación histórica de los riesgos principales del proyecto. Por ejemplo en el caso de PSC sería la valorización de los sobrecostos y los sobreplazos de la fase constructiva de un proyecto (VTRSC).





## Equivalente Cierto en Evaluación de Inversiones de Capital

Con esa valoración se incorpora de manera indirecta la expresión que se resta (al igual que los costos) del flujo de caja bruto para obtener el flujo de caja neto  $F_1$ . Cuando solamente se trabaja con flujos de costos la expresión se puede representar como:

$$VP = \frac{C_1 + \lambda \text{cov}(F_1, r_m)}{1 + r_f} \approx \frac{C_1 + \text{VTRSC}}{1 + r_f}$$

↓  
Técnicas para valorar riesgos = principios de estadística para finanzas

# Probabilidades

□ Interpretación de la probabilidad:

□ Posibilidad de que se produzca un determinado resultado.

□ **Interpretación Objetiva:**

□ Se basa en la frecuencia con que tienen que ocurrir ciertos acontecimientos.

□ **Interpretación Subjetiva:**

□ Se basa en los juicios de valor o en la experiencia de una persona, pero no necesariamente en la frecuencia con que se ha producido realmente un determinado resultado en el pasado:

□ Información diferente o habilidades diferentes para procesar la misma información pueden influir en la probabilidad subjetiva.

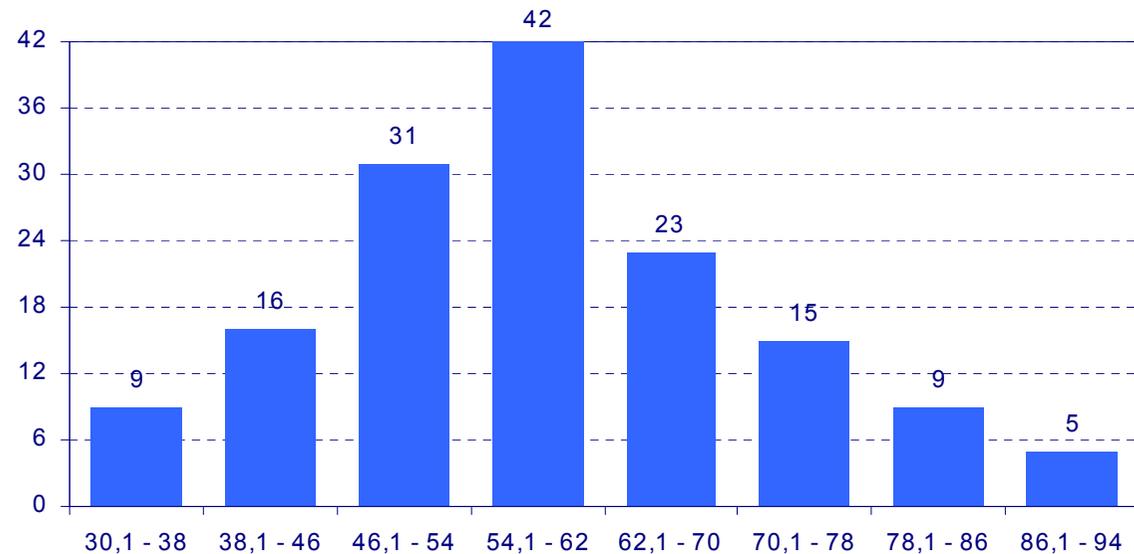




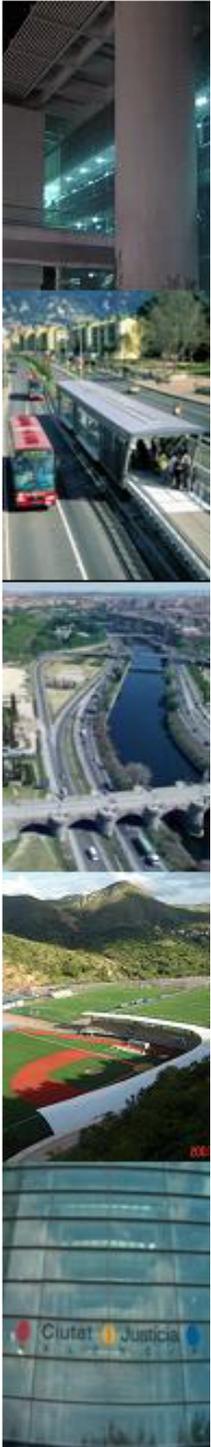
## Descripción de Datos – El Histograma

Un histograma es la representación de la forma como se agrupan una variable en función a su frecuencia (número de veces que se repite un determinado valor de la variable).

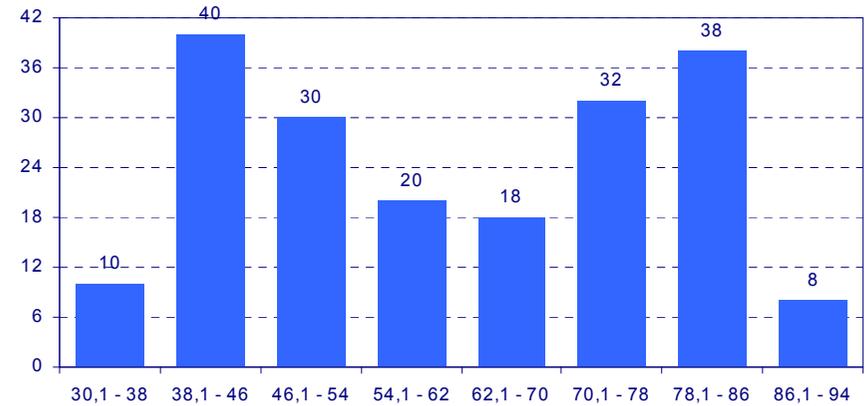
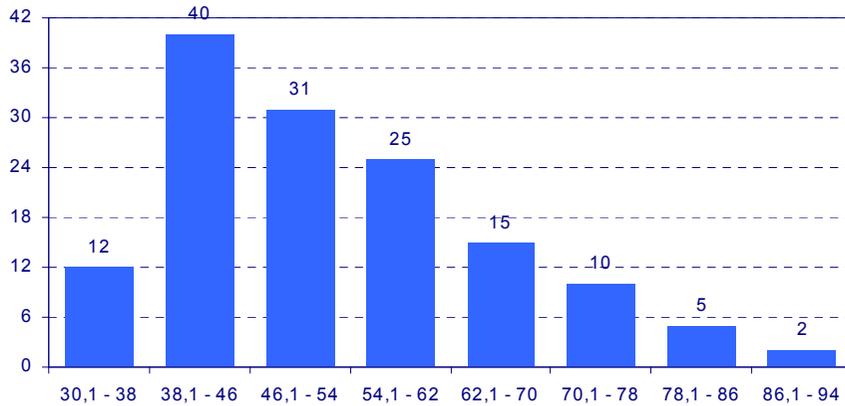
$y_{j-1} - y_j$	$f_j$	$F_j$
30,1 - 38	9	9
38,1 - 46	16	25
46,1 - 54	31	56
54,1 - 62	42	98
62,1 - 70	23	121
70,1 - 78	15	136
78,1 - 86	9	145
86,1 - 94	5	150
$\Sigma$	150	-



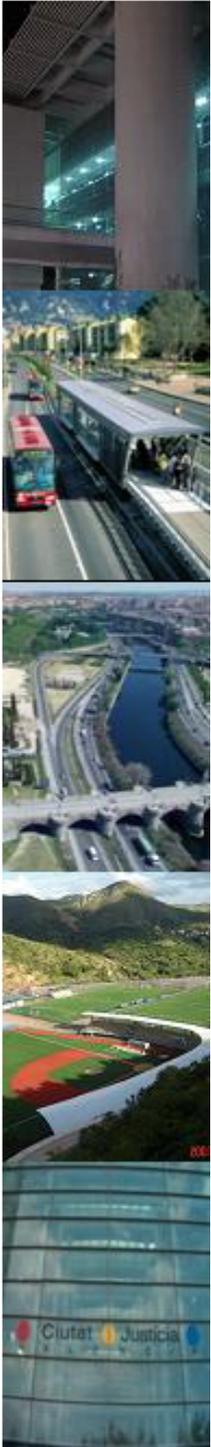
Pero no siempre se agrupan de tal manera que tiende a ser una distribución normal, tal como se ve en los siguientes casos.



# Descripción de Datos – El Histograma



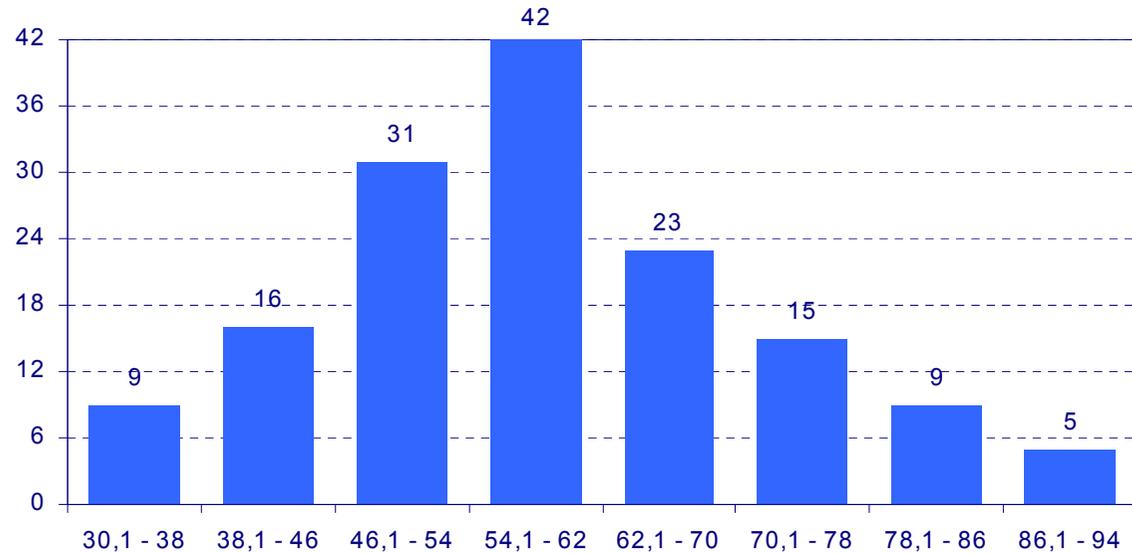
En el primer histograma, la variable tiene distribución asimétrica hacia la derecha, mientras que el segundo histograma se distribuye del tipo bimodal.



## Descripción de Datos – El Histograma

Un histograma es la representación de la forma como se agrupan una variable en función a su frecuencia (número de veces que se repite un determinado valor de la variable).

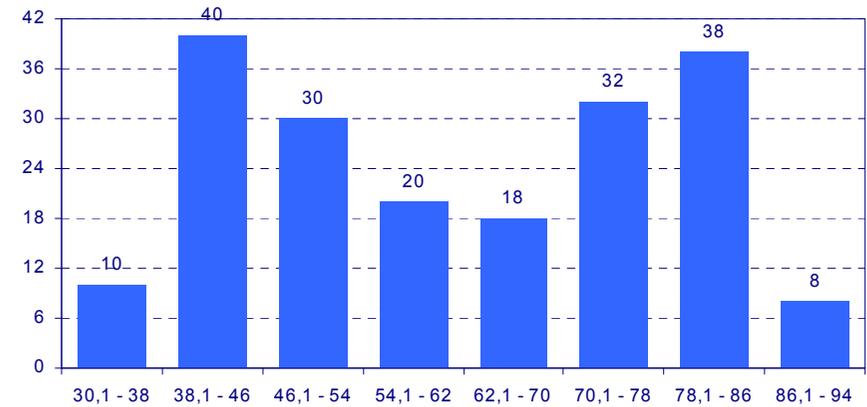
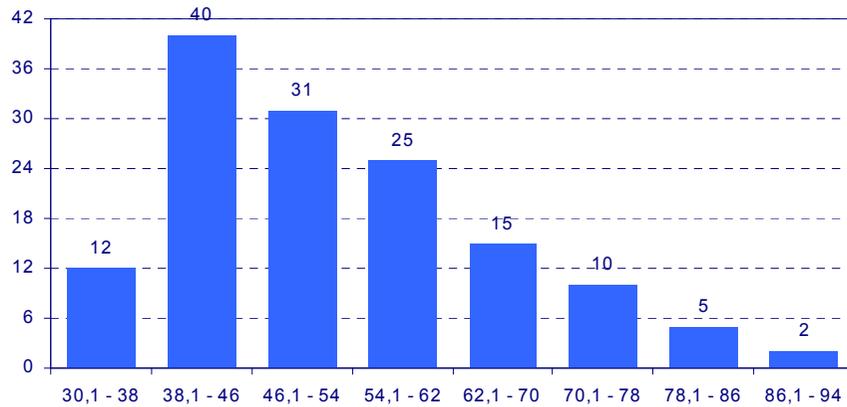
$y_{j-1} - y_j$	$f_j$	$F_j$
30,1 - 38	9	9
38,1 - 46	16	25
46,1 - 54	31	56
54,1 - 62	42	98
62,1 - 70	23	121
70,1 - 78	15	136
78,1 - 86	9	145
86,1 - 94	5	150
$\Sigma$	150	-



Pero no siempre se agrupan de esta manera relativamente simétrica y ordenada, tal como se ve en los siguientes casos.



## Descripción de Datos – El Histograma



En el primer histograma, la variable tiene distribución asimétrica hacia la derecha, mientras que el segundo histograma se distribuye del tipo bimodal. EXCEL CRYSTAL BALL

# Definición del Riesgo

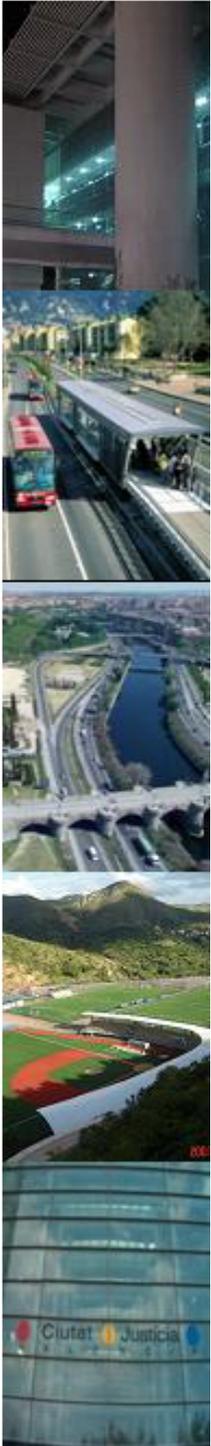
Sea  $X$  una variable aleatoria, una medida de riesgo se define como una función de la variable aleatoria. De manera más precisa, si consideramos que  $X \sim f(X, \theta)$ , i. e., la variable aleatoria  $X$  se distribuye según la función de densidad de probabilidad  $f$  con parámetro  $\theta$ , y además el valor esperado  $E(X)$  y la varianza  $Var(X)$  de la variable aleatoria  $X$ , existen, entonces **una medida del riesgo** es

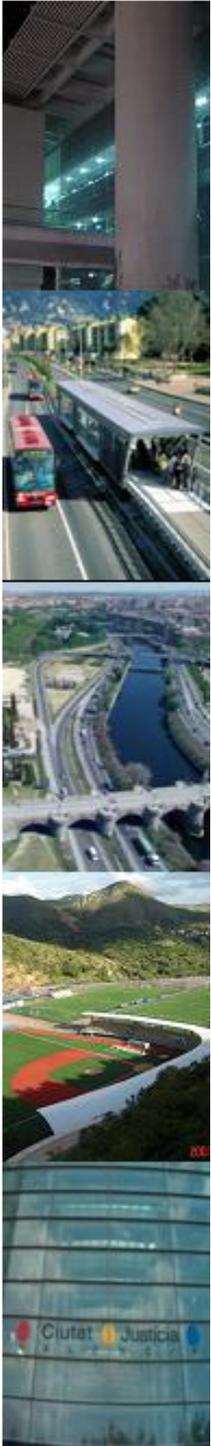
$$\sigma^2 = Var(X)$$



# En finanzas....

- Enfoque de media y varianza o Enfoque riesgo-retorno
- Retornos esperados ( retornos promedios o valores esperados)
- Riesgos basados en variabilidad
- Entonces debemos estudiar el concepto de valor esperado y de varianza



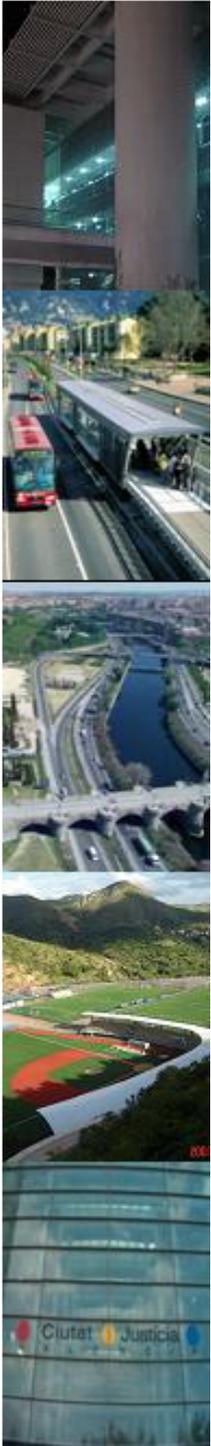


## ❑ Valor esperado

- ❑ Media de los valores correspondientes a todos los resultados posibles ponderada por las probabilidades.
  - ❑ Las probabilidades de cada resultado se utilizan como ponderaciones.
  - ❑ El valor esperado mide la tendencia central, es decir, el rendimiento o el valor que esperamos en promedio.

## ❑ Ejemplo:

- ❑ Inversión en prospecciones petrolíferas:
  - ❑ Dos resultados posibles:
    - ❑ Éxito: el precio crece de 30 dólares a 40 por acción.
    - ❑ Fracaso: el precio cae de 30 dólares a 20 por acción.



## □ Ejemplo:

□ Probabilidad objetiva:

□ 100 exploraciones, 25 éxitos y 75 fallos.

□ Probabilidad de éxito ( $Pr$ ) =  $1/4$  y la probabilidad de fracaso =  $3/4$ .

## □ Valor Esperado

**Valor**

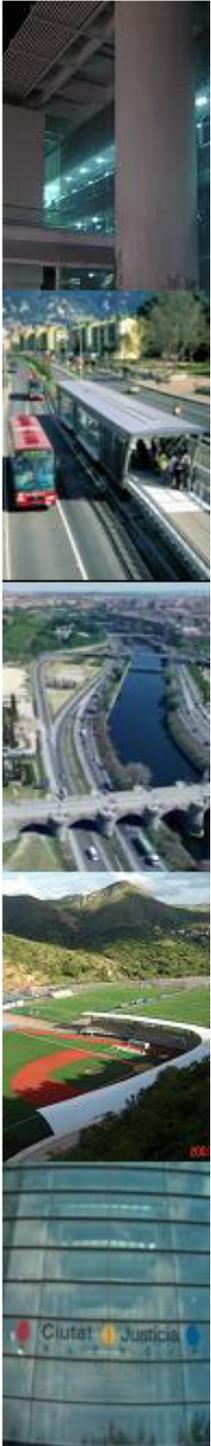
$$\text{Esperado} = Pr(\text{Exito}) \times (40\$ / \text{Acción}) + Pr(\text{Fracaso}) \times (20\$ / \text{Acción})$$

**Valor**

$$\text{Esperado} = \frac{1}{4} \times (40\$ / \text{Acción}) + \frac{3}{4} \times (20\$ / \text{Acción})$$

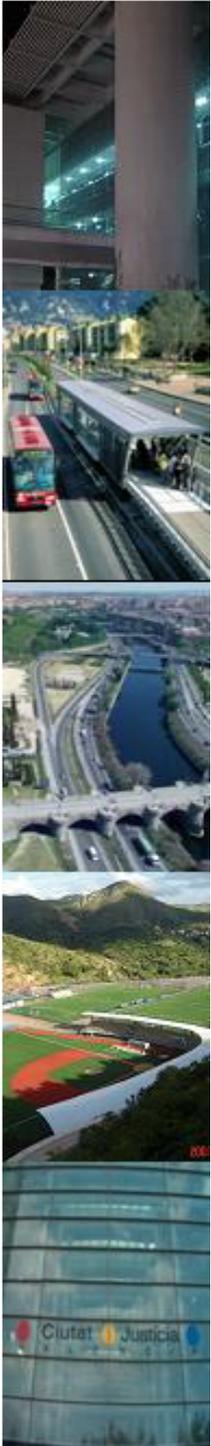
**Valor**

$$\text{Esperado} = 25\$ / \text{Acción}$$



- Si hay dos resultados posibles que tiene unos rendimientos de  $X_1$  y  $X_2$ .
- Las probabilidades de cada resultado vienen dadas por  $Pr_1$  y  $Pr_2$ .
- En términos generales, el valor esperado es:

$$E[X] = Pr_1 X_1 + Pr_2 X_2 + \dots + Pr_n X_n$$

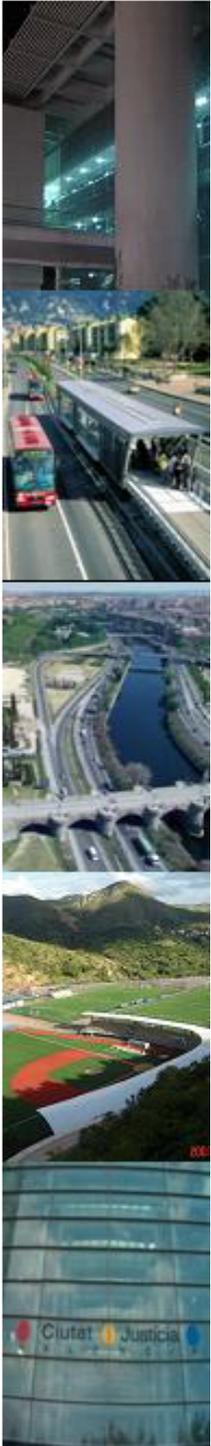


- ❑ **Variabilidad ó Volatilidad**
- ❑ Grado en que pueden variar los posibles resultados de un acontecimiento incierto.
- ❑ **Ejemplo:**
  - ❑ Supongamos que estamos eligiendo entre dos puestos de trabajo en el área de ventas que tienen la misma renta esperada (1.500 dólares).
    1. El primero se basa enteramente en comisiones por venta.
    2. El segundo es asalariado con valor fijo sin contingencias.

# Ejemplo riesgos en W/P

## □ Ejemplo

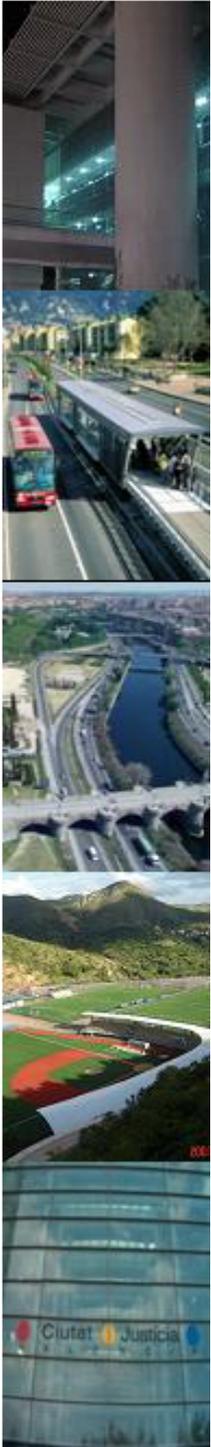
1. El primer trabajo tiene dos rendimientos iguales probables: US\$ 2,000, en el caso de que las ventas sean cuantiosas y US\$ 1,000, en el caso de que las ventas sean menores.
2. En el segundo es muy probable (una probabilidad de 0.99) que ganemos US\$ 1,510, pero hay una posibilidad de 0.01 de que la compañía quiebre, en cuyo caso percibiríamos una indemnización por desempleo de US\$ 510



# La Descripción del Riesgo

## La Renta de los Empleos de Ventas

	Resultado 1		Resultado 2		Renta Esperada
	Probabilidad	Renta (\$)	Probabilidad	Renta (\$)	
E1: Comisión	0.50	2,000	0.50	1,000	1,500
E2: Sueldo Fijo	0.99	1,510	0.01	510	1,500



□ La renta de los empleos de ventas

□ Renta esperada del Empleo 1:

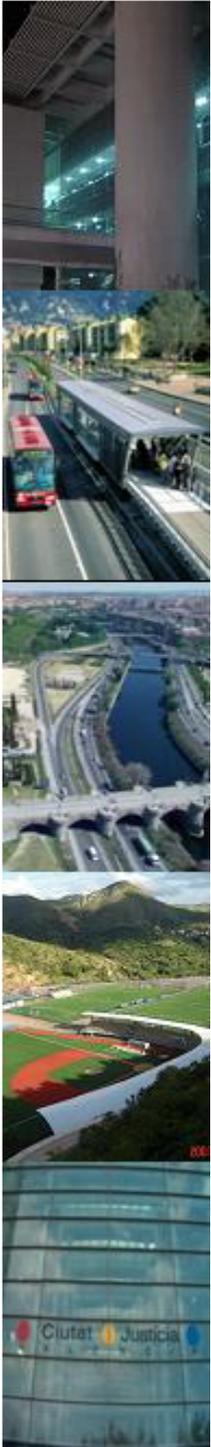
$$E[X_1] = 0.5 \times (\text{US\$ } 2,000) + 0.5 \times (\text{US\$ } 1,000) = \text{US\$ } 1,500$$

□ Renta esperada del Empleo 2:

$$E[X_1] = 0.99 \times (\text{US\$ } 1,510) + 0.01 \times (\text{US\$ } 510) = \text{US\$ } 1,500$$

□ Mientras que los valores esperados son iguales, la variabilidad no lo es.

□ **Cuanto mayor sea la variabilidad de los valores esperados, mayor riesgo.**



## ❑ Desviación:

- ❑ Diferencia entre el rendimiento esperado y el real.

Desviaciones con respecto a la Renta Esperada (\$)				
	Resultado 1	Desviación	Resultado 2	Desviación
Empleo 1	2,000	500	1,000	-500
Empleo 2	1,510	10	510	-990

# Variabilidad

## □ Variabilidad:

- Las diferencias negativas deben corregirse.
- La **desviación típica** mide la raíz cuadrada de la media del cuadrado de las desviaciones de los resultados con respecto a su valor esperado.
- La ecuación de la desviación típica es la siguiente:

$$\sigma = \sqrt{Pr_1 [X_1 - E[X]]^2 + Pr_2 [X_2 - E[X]]^2}$$

# Variabilidad

Cálculo de la Varianza (\$)

	Resultado 1	Cuadrados de las desviaciones	Resultado 2	Cuadrado de las desviaciones	Desviación media al cuadrado	Desviación típica
Empleo 1	2,000	250,000	1,000	250,000	250,000	500
Empleo 2	1,510	100	510	980,100	9,900	99.50

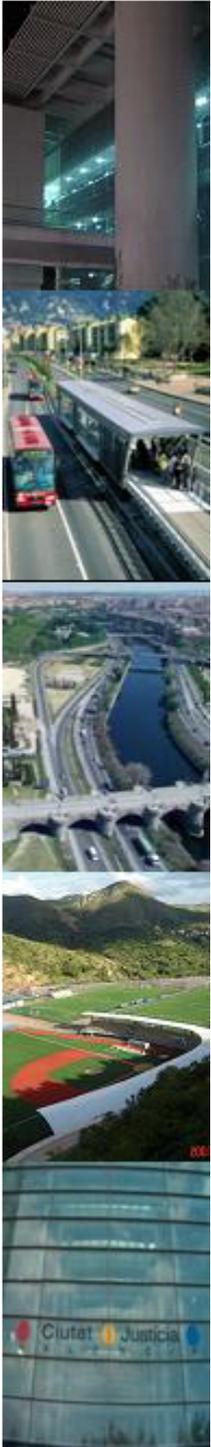
Las desviaciones típicas de los dos empleos son:

$$\sigma_1 = \sqrt{0.5 \times 250,000 + 0.5 \times 250,000} = \sqrt{250,000}$$

→  $\sigma_1 = 500$  **Mayor Riesgo**

$$\sigma_2 = \sqrt{0.99 \times 100 + 0.01 \times 980,100} = \sqrt{9,900}$$

→  $\sigma_2 = 99,50$



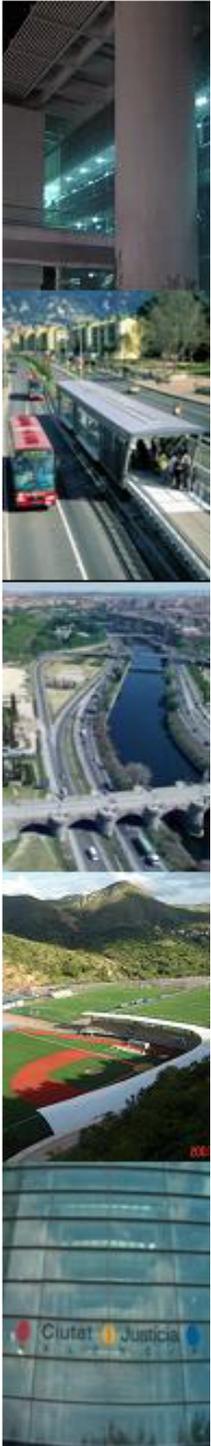
# Definiciones Formales



# Valor Esperado

El valor esperado de una variable aleatoria  $X$ , el cual se denota como  $E(X)$ , es el valor medio del resultado de un gran número de ensayos experimentales.

$$E(X) = \mu = \begin{cases} \sum_i x_i \cdot f(x_i), & \text{para el caso discreto} \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx, & \text{para el caso continuo} \end{cases}$$

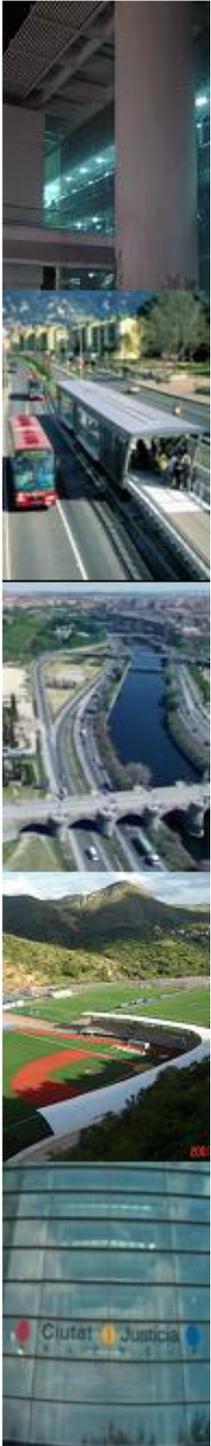


# Varianza

Es una medida que, en promedio, cuantifica el nivel de dispersión o de variabilidad de los valores de una variable cuantitativa con respecto a su valor esperado.

$$\text{Var}(X) = E(X - E(X))^2 = \begin{cases} \sum_i (x_i - \mu)^2 \cdot f(x_i), & X \text{ discreto} \\ \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx, & X \text{ continuo} \end{cases}$$

$$\text{Var}(X) = E(X - E(X))^2 = E(X^2) - E(X)^2$$

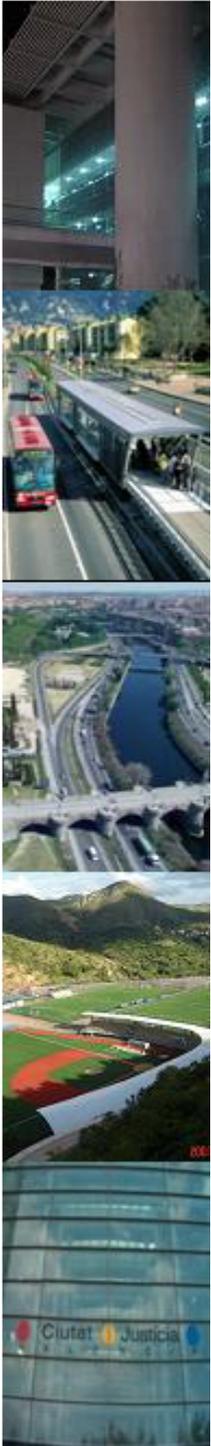


# Volatilidad

La volatilidad es un indicador que cuantifica con que probabilidad se producen los cambios de la variable aleatoria en torno su valor esperado.

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$$

La definición de la volatilidad se ve reflejado en la expresión de la izquierda, pero para el cálculo de la volatilidad histórica se emplea la expresión de la derecha.



# Cálculo de la Volatilidad Histórica

Inversión	Rendimiento Anual Promedio
Acciones pequeñas	22.11%
S&P 500	11.40%
Bonos corporativos	6.52%
Títulos del Tesoro	3.87%



# Tabla de Retornos

Año	S&P 500 Índice		S&P 500 retornos		3 – Month T – Bill Retornos
1995	615.93				
1996	740.74		23.0%		5.1%
1997	970.43		33.4%		5.2%
1998	1,229.23		28.6%		4.9%
1999	1,469.25		21.0%		4.8%
2000	1,320.28		-9.1%		6.0%
2001	1,148.08		-11.9%		3.3%
2002	897.82		-22.1%		1.6%
2003	1,111.92		28.7%		1.0%
2004	1,211.92		10.9%		1.4%



# Cálculo de la Volatilidad Histórica

Con los datos de la Tabla anterior, mencione cuáles son la varianza y volatilidad de los rendimientos del S&P 500 para los años 1996 a 2004.

El rendimiento anual promedio del S&P 500 durante este periodo había sido de **11.4%** . Por lo tanto,

$$\begin{aligned}\text{Var}(R) &= \frac{1}{T-1} \sum_t (R_t - \bar{R})^2 \\ &= \frac{1}{9-1} \left[ (0.230 - 0.114)^2 + (0.334 - 0.114)^2 + \dots + (0.109 - 0.114)^2 \right] \\ &= 4.24\%\end{aligned}$$

# Cálculo de la Volatilidad Histórica

Entonces, la volatilidad o desviación estándar es

$$SD(R) = \sqrt{\text{Var}(R)} = \sqrt{0.0424} = 20.6\%$$

La desviación estándar de los rendimientos se calcula para cuantificar las diferencias en los riesgos de cada activo financiero

Inversión	Volatilidad
Acciones pequeñas	42.75%
S & P 500	20.60%
Bonos Corporativos	7.17%
Títulos del Tesoro	3.18%

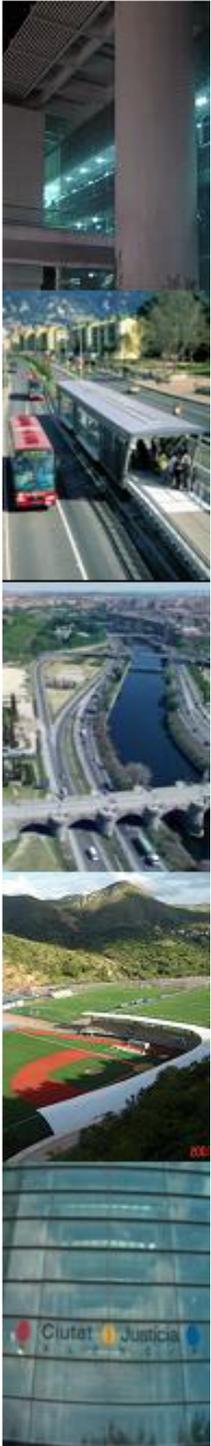
# Definición del Riesgo

Sea  $x$  una variable aleatoria, una medida de riesgo se define como una función de la variable aleatoria.

$$X \sim f(X, \theta)$$

Entonces **una medida del riesgo** es

$$\sigma^2 = \text{Var}(X)$$

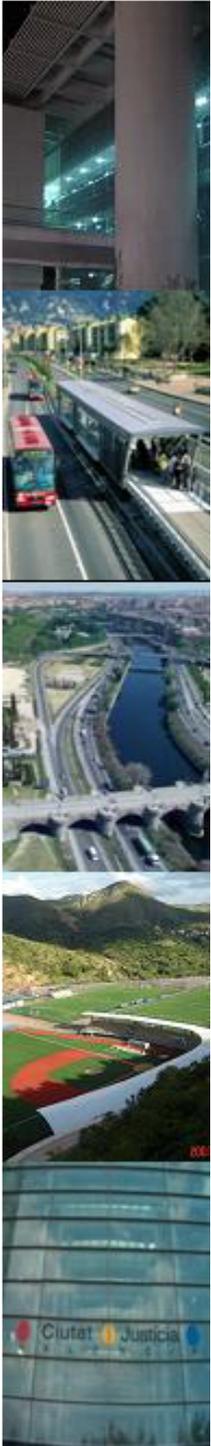


# Definición del Riesgo

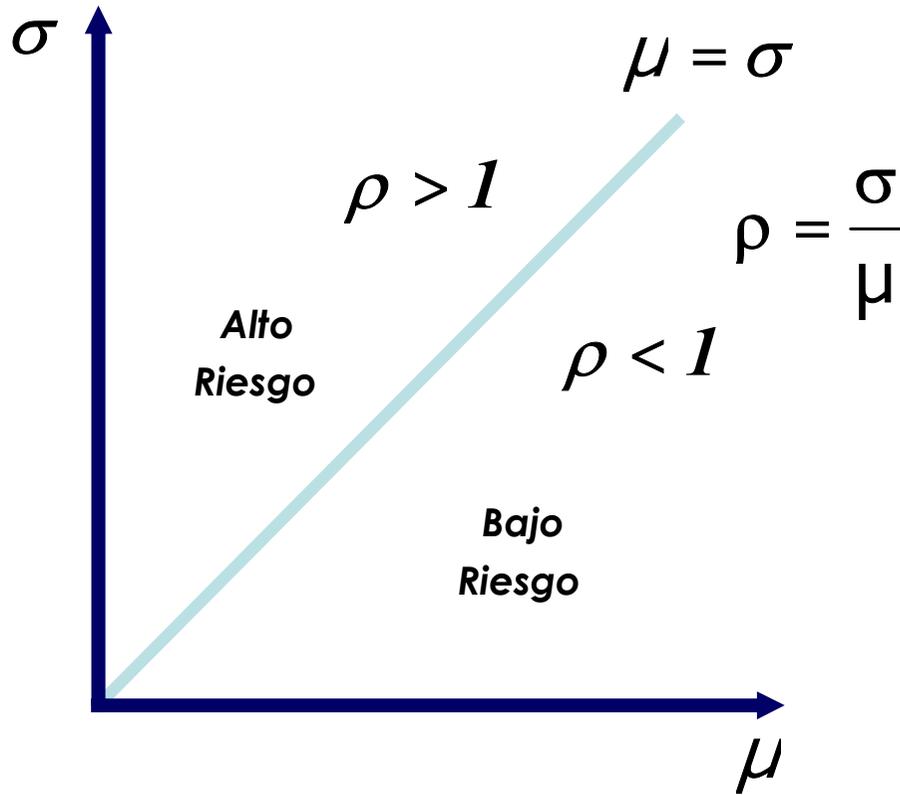
Una medida alternativa para medir la variabilidad del riesgo, viene dado por el Coeficiente de Variación, que es el cociente entre la volatilidad y su valor esperado,

$$\rho = \frac{\sigma}{\mu}$$

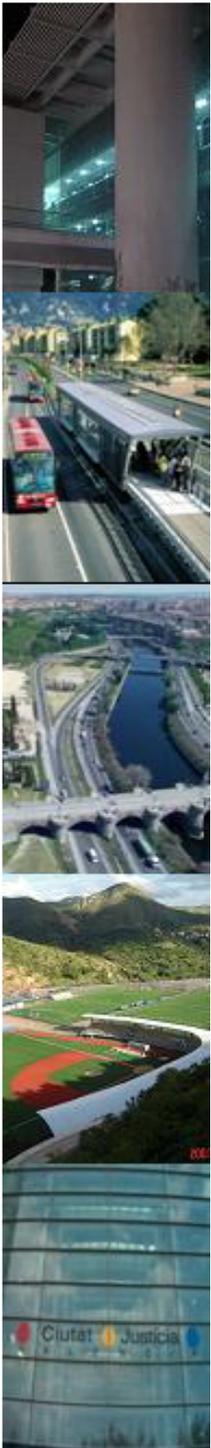
Donde  $\sigma = \sqrt{\text{Var}(X)}$  y  $\mu = E(X)$  .

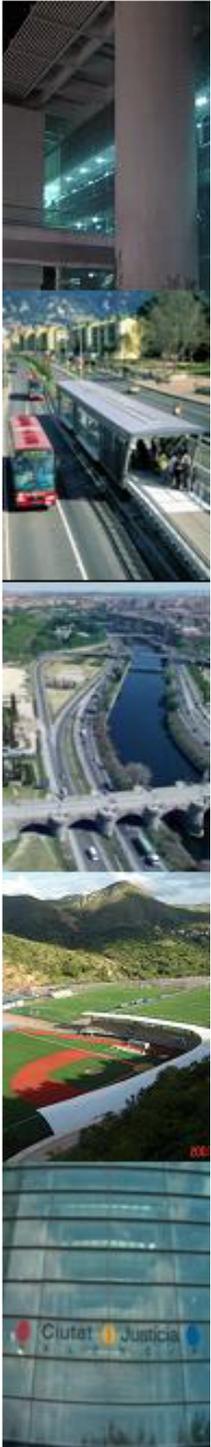


# Definición del Riesgo



Si el CV  $\rho$  es muy bajo (empíricamente se puede considerar como bajos valores de  $\rho < 30\%$ ) tienen asociadas distribuciones cuyas varianzas son bajas y viceversa;  $E(X) = \mu$  donde es fija y conocida



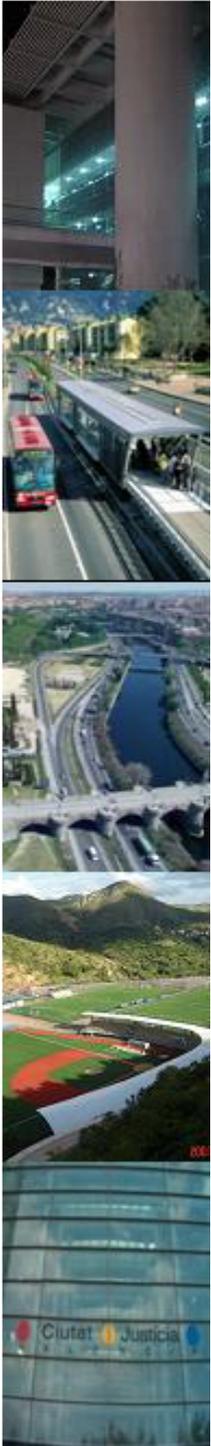


# Supuestos fundamentales para el tratamiento del riesgo en finanzas



# Supuestos Fundamentales para la Incorporación del Riesgo en la Valorización de Activos bajo enfoque media-varianza

1. Los inversionistas son aversos al riesgo y maximizan la utilidad esperada bajo el enfoque de VNM
2. la varianza (o desviación estándar) es una medida del riesgo
3. Los retornos tienen una distribución normal y/o la función de utilidad es cuadrática
  - Maximizar Retorno s.a. Volatilidad
  - Minimizar Volatilidad s.a. Retorno



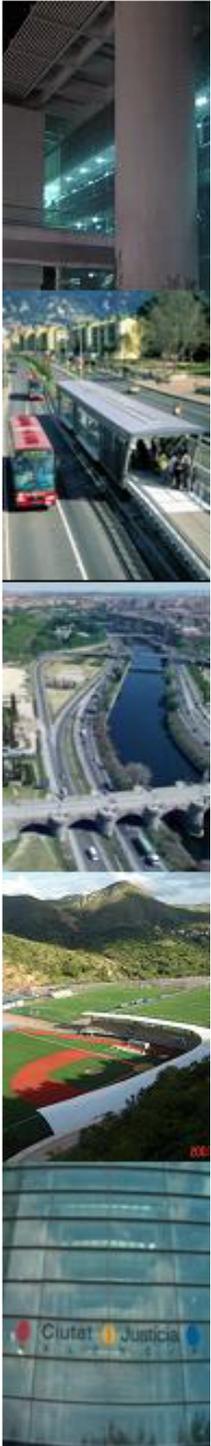
# Enfoque media varianza

- ❑ Solo es posible justificar, y en consecuencia trabajar en un mundo riesgo-retorno si:
  - ❑ Los inversionistas maximizan la utilidad esperada
  - ❑ La función de utilidad es cuadrática y/o los retornos siguen una función de distribución Normal



# Observación

- ❑ Hay juegos que tienen valor esperado infinito y sin embargo no hay disposición a pagar una suma finita por ingresar. (**Paradoja de San Petersburgo**)
- ❑ Hay juegos que tienen valor esperado negativo y sin embargo hay personas que pagan una suma finita por ingresar. (**Juegos de Azar como la lotería u otros**).



# Constatación

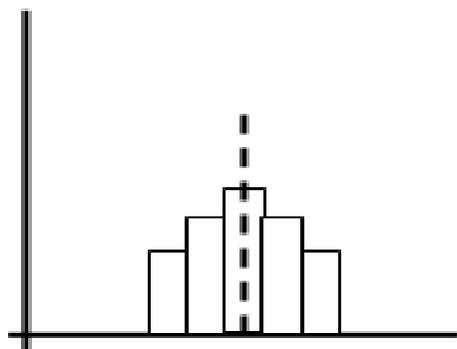
- ❑ Los individuos no maximizan el Valor Esperado de la riqueza.
- ❑ Los individuos maximizan la Utilidad Esperada de la riqueza
- ❑ Teoría de la Utilidad Esperada de Von Neumann Morgenstern

$$E[U(W)] = \sum_i U(W_i) \times p_i$$

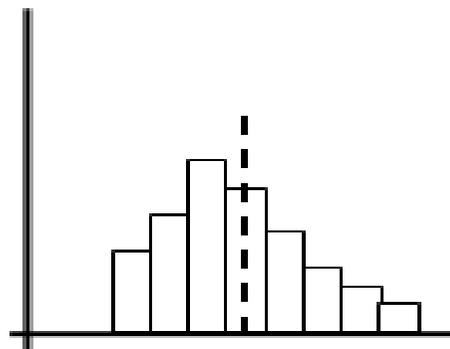
$$E[U(W)] = \int U(W) \times f_i \times dW$$



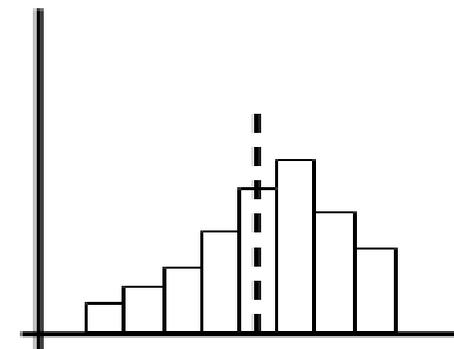
# Distribución Normal de los Retornos Bajo TCL



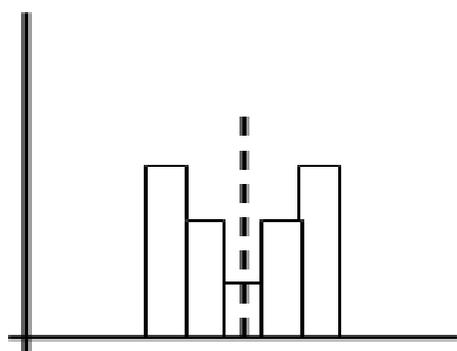
$\bar{x}$   
**Simétrica**



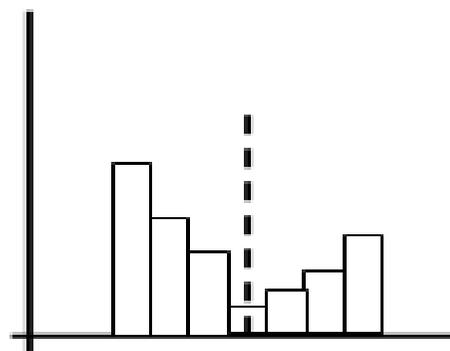
$\bar{x}$   
**Asimétrica a la derecha**



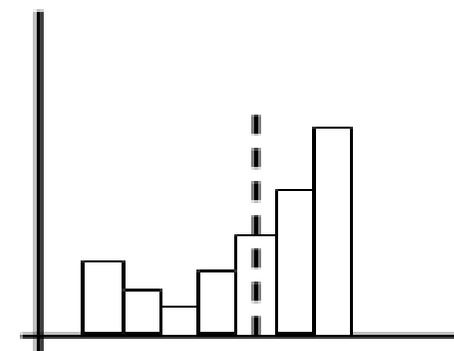
$\bar{x}$   
**Asimétrica a la izquierda**



$\bar{x}$   
**Simétrica**



$\bar{x}$   
**Asimétrica a la derecha**



$\bar{x}$   
**Asimétrica a la izquierda**

# Utilidad Esperada Bajo Función Cuadrática

Función cuadrática

$$U(W) = W - bW^2 \quad \text{con } b > 0$$

$$U'(W) = 1 - 2bW \quad \text{Utilidad Marginal Positiva}$$

$$U''(W) = -2b < 0 \quad \text{Utilidad Marginal Decreciente}$$

Tomando valor esperado a la primera ecuación:

$$E[U(W)] = E(W) - bE(W^2)$$

$$E[U(W)] = \mu - bE(W^2)$$

Sabemos que por definición la varianza es igual a:

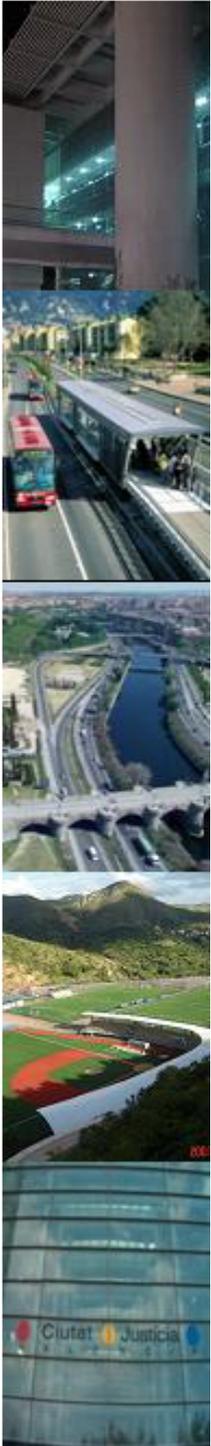
$$\sigma_W^2 = E(W^2) - [E(W)]^2$$

Sustituyendo en 1a se obtiene

$$E[U(W)] = \mu - b(\mu^2 + \sigma_W^2) = U(\mu, \sigma)$$

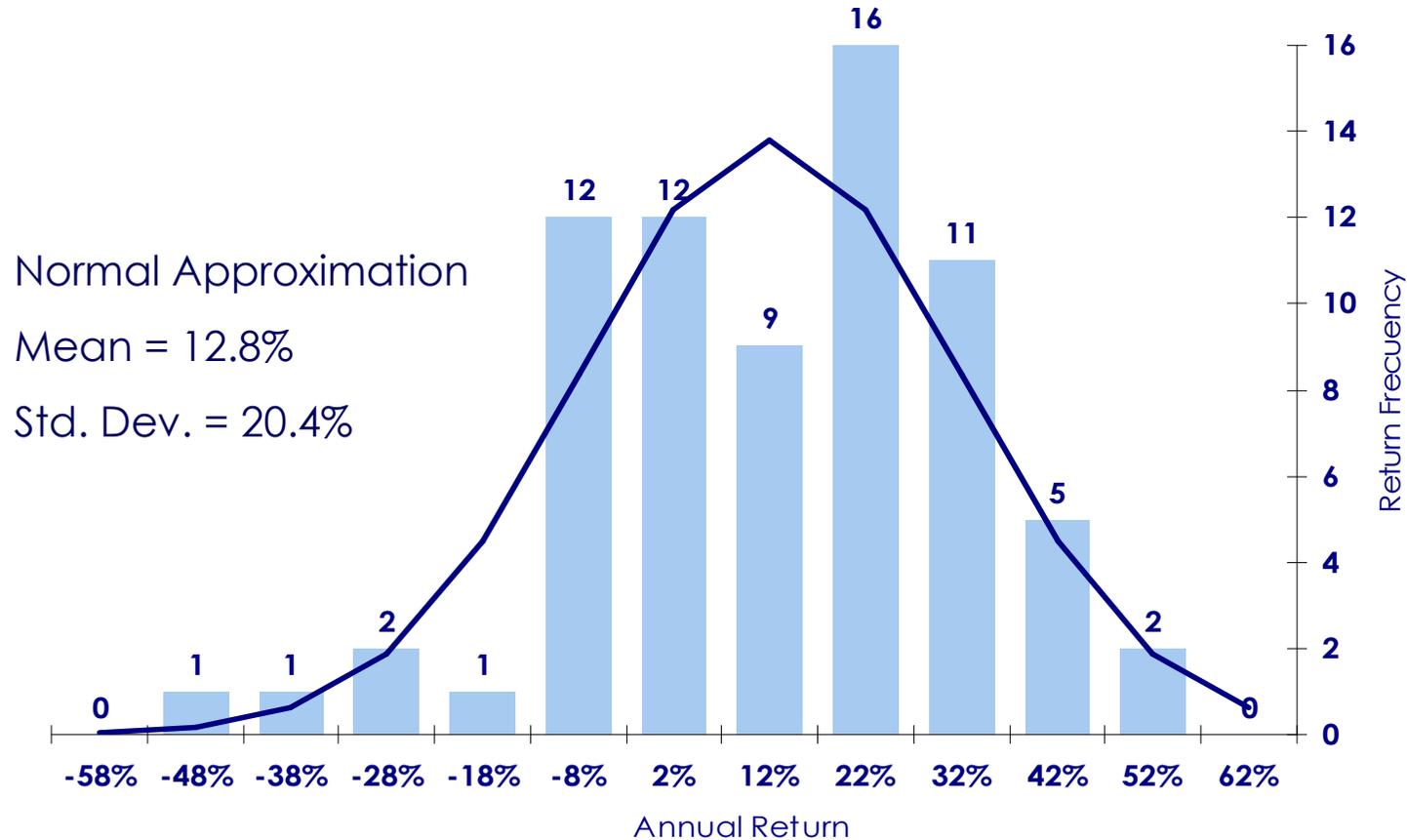
# Riesgos en un mundo de funciones de distribución Normal

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2}$$



# Frecuencias de los Retornos

S & P 500 Return Frequencies



Source: © *Stocks, Bonds, Bills, and Inflation 2002 Yearbook™*, Ibbotson Associates, Inc., Chicago (annually updates work by Roger G. Ibbotson and Rex A. Sinquefeld). All rights reserved.





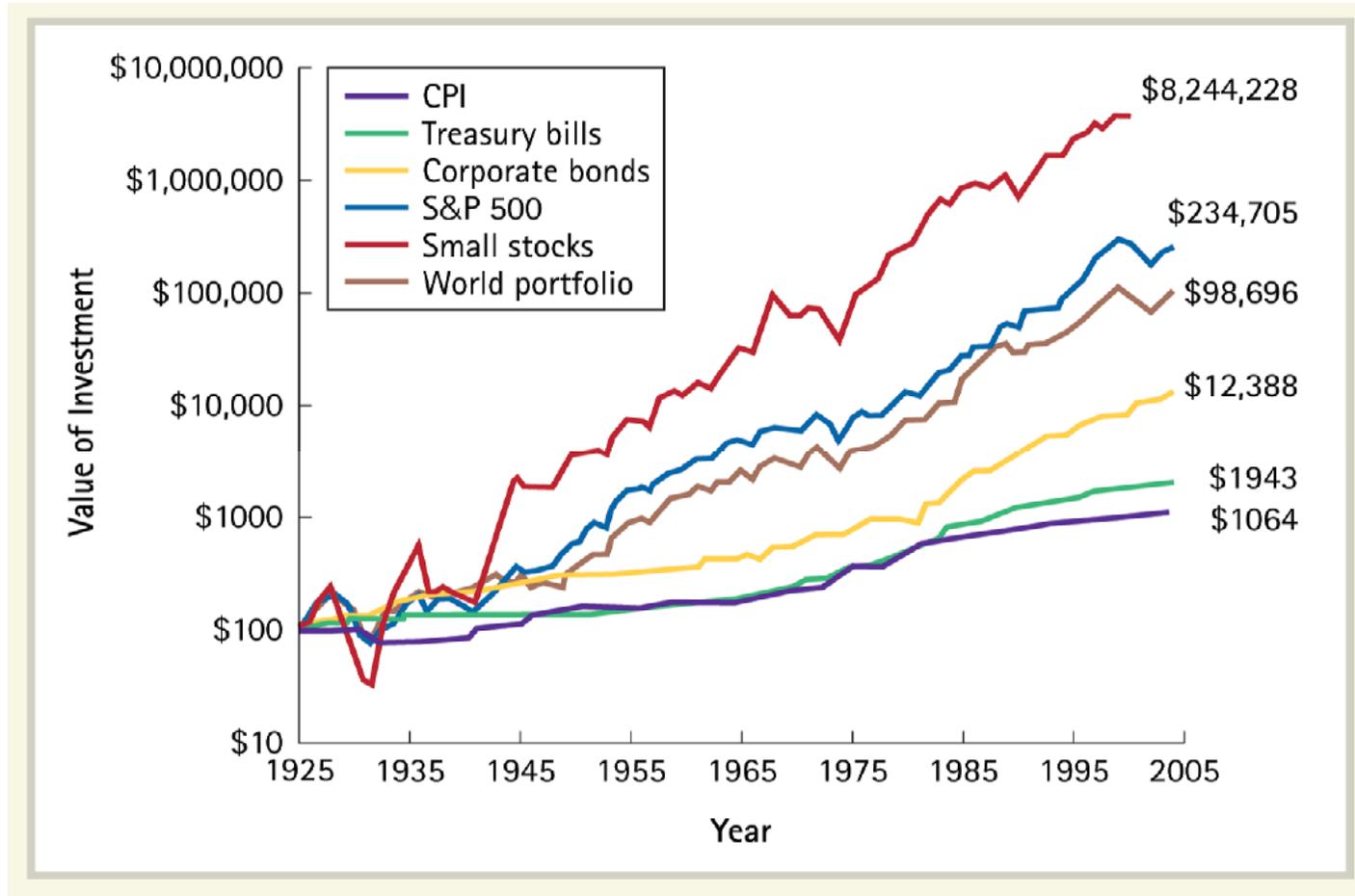
# Rendimientos anuales promedio para acciones pequeñas, acciones grandes (S&P 500), bonos corporativos, y títulos del Tesoro, de los Estados Unidos, 1926 – 2004

Inversión	Rendimiento Anual Promedio
Acciones pequeñas	22.11%
S&P 500	12.32%
Bonos corporativos	6.52%
Títulos del Tesoro	3.87%

Investment	Return Volatility (Standard Deviation)
Small stock	42.75%
S&P 500	20.36%
Corporate bonds	7.17%
Treasury bill	3.18%

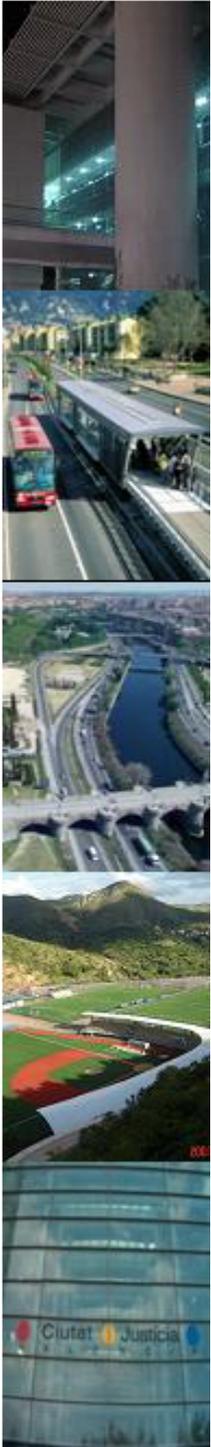


# US\$ 100 invertidos en 1925 en distintos instrumentos financieros



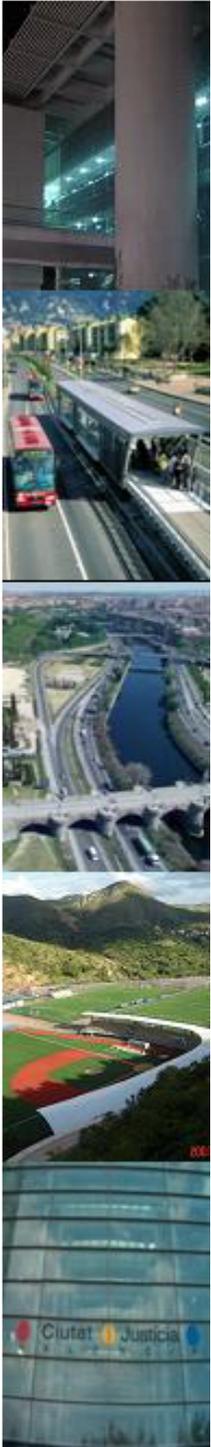
Source: Chicago Center for Research in Security Prices (CRSP) for U.S. stocks and CPI, Global Finance Data for the World Index, Treasury bills and corporate bonds citado en Berk y DeMarzo (2008)





## Test de Normalidad

$$x_1, x_2, \Lambda, x_n$$

## Test de Normalidad

De tal manera que se representa los siguientes pares ordenados:

$$(x_1, F(x_1)), (x_2, F(x_2)), \Lambda, (x_n, F(x_n))$$

$$(x_1, F_n(x_1)), (x_2, F_n(x_2)), \Lambda, (x_n, F_n(x_n))$$

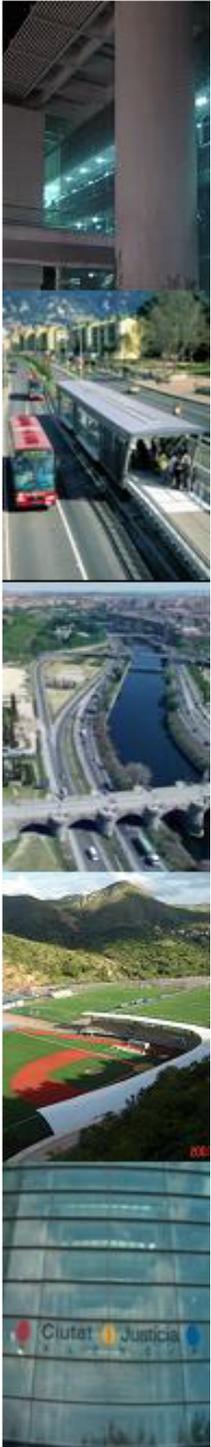
Donde  $F_n$  representa la función de distribución acumulada del conjunto de datos y  $F$  la función de distribución acumulada esperada (por ejemplo, la distribución normal). Ambas curvas son representadas en un mismo gráfico de tal manera que los puntos

$$(x_1, F(x_1)), (x_2, F(x_2)), \Lambda, (x_n, F(x_n))$$

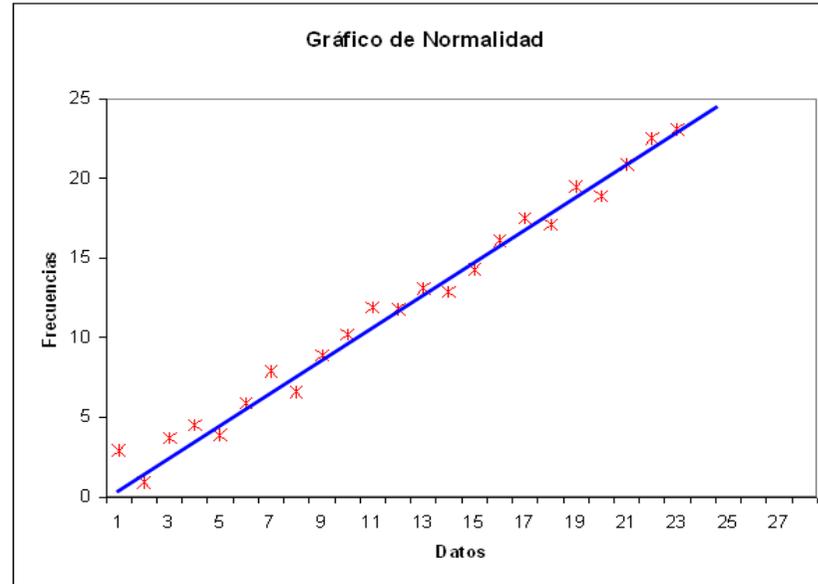
se encuentran sobre la recta  $y=x$ . Por consiguiente aceptaremos la hipótesis de normalidad siempre que los puntos

$$(x_1, F_n(x_1)), (x_2, F_n(x_2)), \Lambda, (x_n, F_n(x_n))$$

se encuentren próximos a la recta.

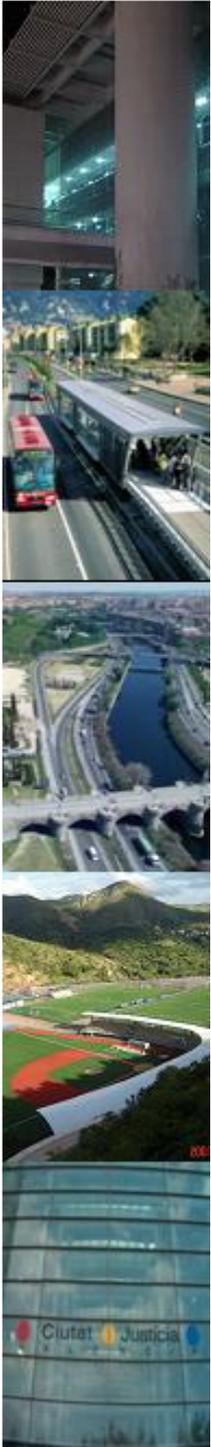


# Test de Normalidad



Muchos métodos no paramétricos han sido creados para medir la bondad del ajuste de una distribución normal, con la finalidad de contrastar si una muestra sigue una determinada función de distribución (no solo la normal). A continuación se presentan los más usados:



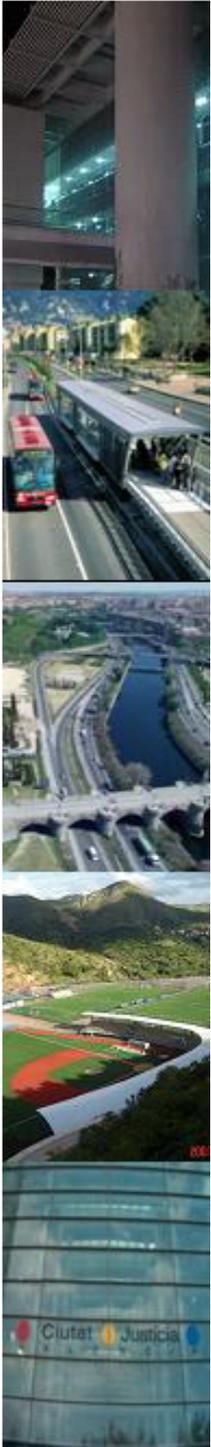


## Test de Kolmogorov - Smirnov

Esta prueba se aplica solo para variables continuas, y se utiliza para comprobar la hipótesis nula de que la muestra procede de una distribución normal. Se fundamenta en la comparación de la función de distribución acumulada de los datos observados, con respecto a la función de distribución esperada, midiendo la máxima distancia entre ambas curvas, que no deberá exceder un valor crítico, que se obtiene de una tabla de probabilidad. Es decir,

$$D = \text{máx} \{F_n(x_i) - F(x_i)\}$$

La hipótesis nula se acepta cuando el ~~D~~ observado es inferior al  $D$  esperado, que se encuentra en la tabla de la prueba de una muestra de Kolmogorov – Smirnov.

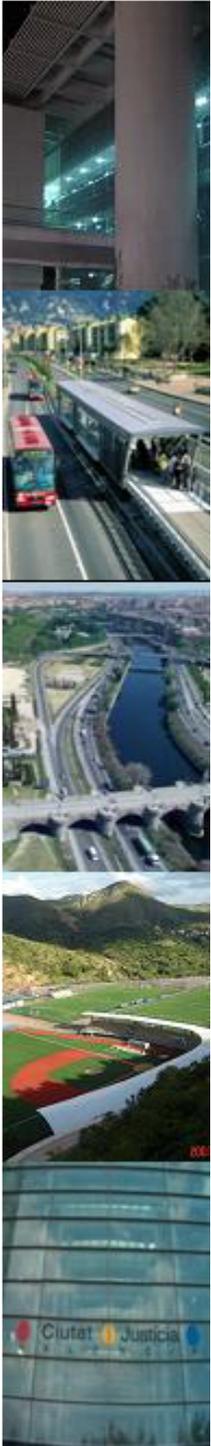


## Test de Shapiro – Wild

Esta prueba de normalidad es recomendable, cuando se cuente con muestras reducidas  $n < 30$ . Este estadístico mide como los datos observados se ajusta a la recta (Recta probabilística normal, recta de  $45^\circ$ ) y no a la distancia a la distribución normal. Este estadístico se formula de la siguiente manera:

$$W = \frac{1}{\sum_{j=1}^n (x_j - \mu)^2} \left[ \sum_{j=1}^h (x_{j,n} (x_{(n-j+1)} - x_j)) \right]^2$$

Donde,  $n$  es el número de datos,  $x_j$  es el dato en orden ascendente de la muestra que ocupa el lugar  $j$ ,  $\mu$  es la media,  $h$  es  $n/2$  si  $n$  es par, o  $(n-1)/2$  si  $n$  es impar  $x_{j,n}$  es un valor tabulado. La hipótesis nula se acepta cuando el valor de  $W$  es superior al valor de ajuste tabulado



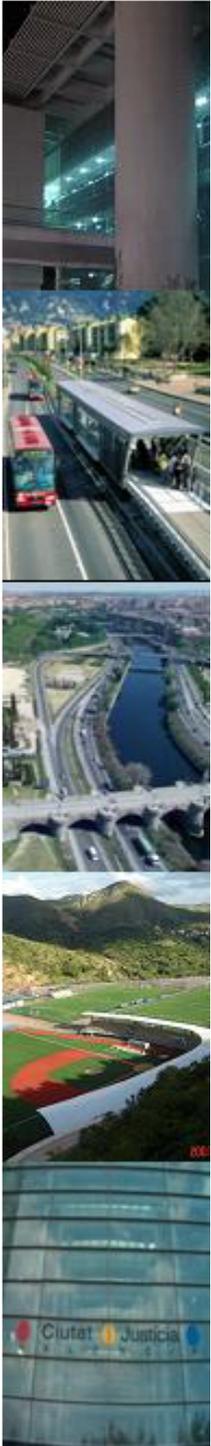
## Test de Chi – Cuadrado

Con un cierto grado de confianza previamente establecido, nos permite determinar si los datos  $x_1, x_2, \dots, x_n$  están asociados con la distribución normal, mediante la siguiente expresión:

$$\chi^2 = \sum_{j=1}^n \frac{(F_n(x_i) - F(x_i))^2}{F(x_i)}$$

Si el valor del estadístico es mayor que el valor en las tablas de la distribución Chi – Cuadrado con  $(f-1)(c-1)$  grados de libertad, al nivel de significación  $\alpha$ , se rechaza la hipótesis nula y se acepta que existe relación entre ambas variables.

Para poder realizar la prueba es necesario que las frecuencias teóricas de las casillas sean 5 como mínimo y el tamaño de la muestra mínimo es de 50. (Grande, 2007)

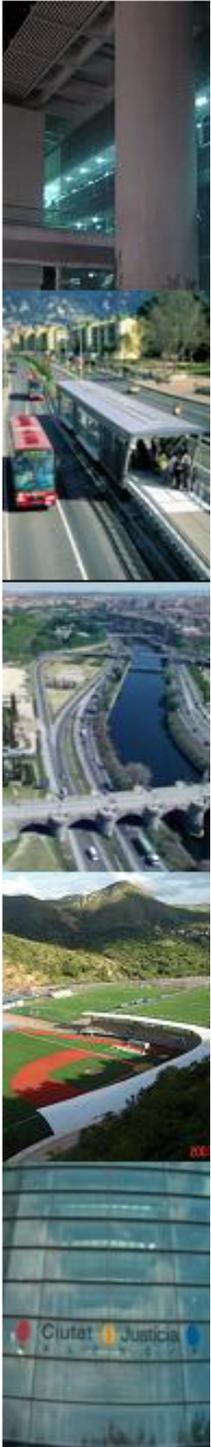


## Test de Jarque Vera

Analiza la normalidad o no normalidad de un conjunto de datos, comparando la diferencia entre los coeficientes de asimetría y curtosis de las observaciones, con respecto a una distribución normal. Esta es una prueba asintótica, o de grandes muestras. La expresión de este estadístico es el siguiente:

$$JB = n \times \left[ \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right] \sim \chi_2^2$$

Siendo  $S$  la asimetría y  $K$  la curtosis. Bajo la hipótesis nula de distribución normal, el estadístico Jarque Bera se distribuye como una Chi – cuadrado con 2 grados de Libertad.



## Test de Jarque Vera

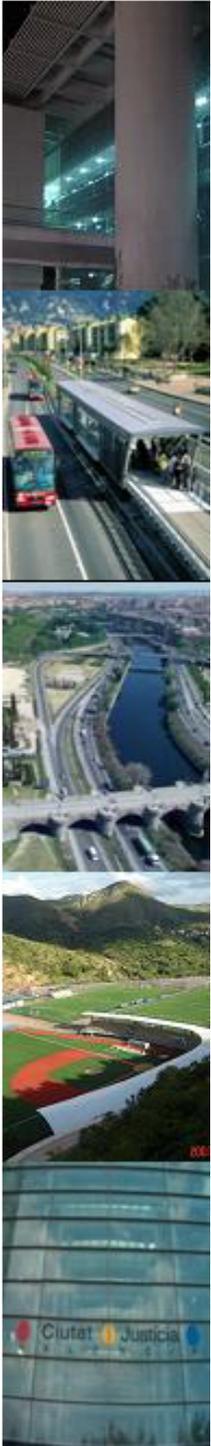
Este test contrasta la siguiente hipótesis nula: *La distribución es normal*, de las tablas correspondientes tenemos que con un 95% de confianza se tiene el valor de 5.99, es decir, se puede concluir que para valores menores a 6 del **JB** no se rechaza el supuesto de normalidad.

Por ejemplo, supongamos que contamos con una muestra de 97 sobrecostos del sector de infraestructura, los cuales tienen los siguientes estadísticos:

$$\text{Simetría} = 0.149 \quad \text{Kurtosis} = 2.942 \quad \Rightarrow \quad \text{JB} = 0.374$$

Esto nos permite aceptar la hipótesis de nula de la normalidad, con un grado de confianza superior al 99%.





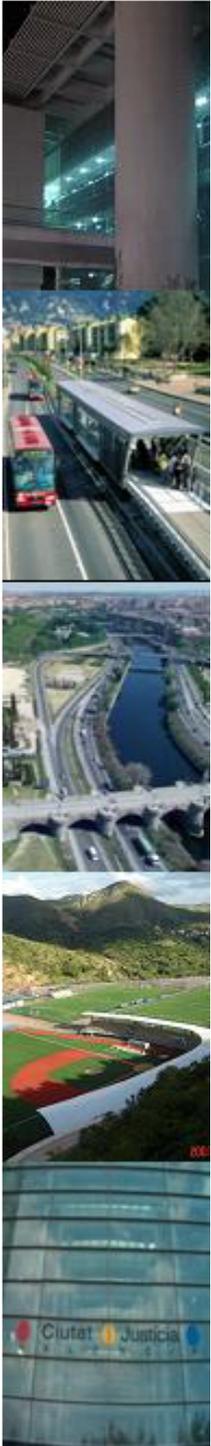
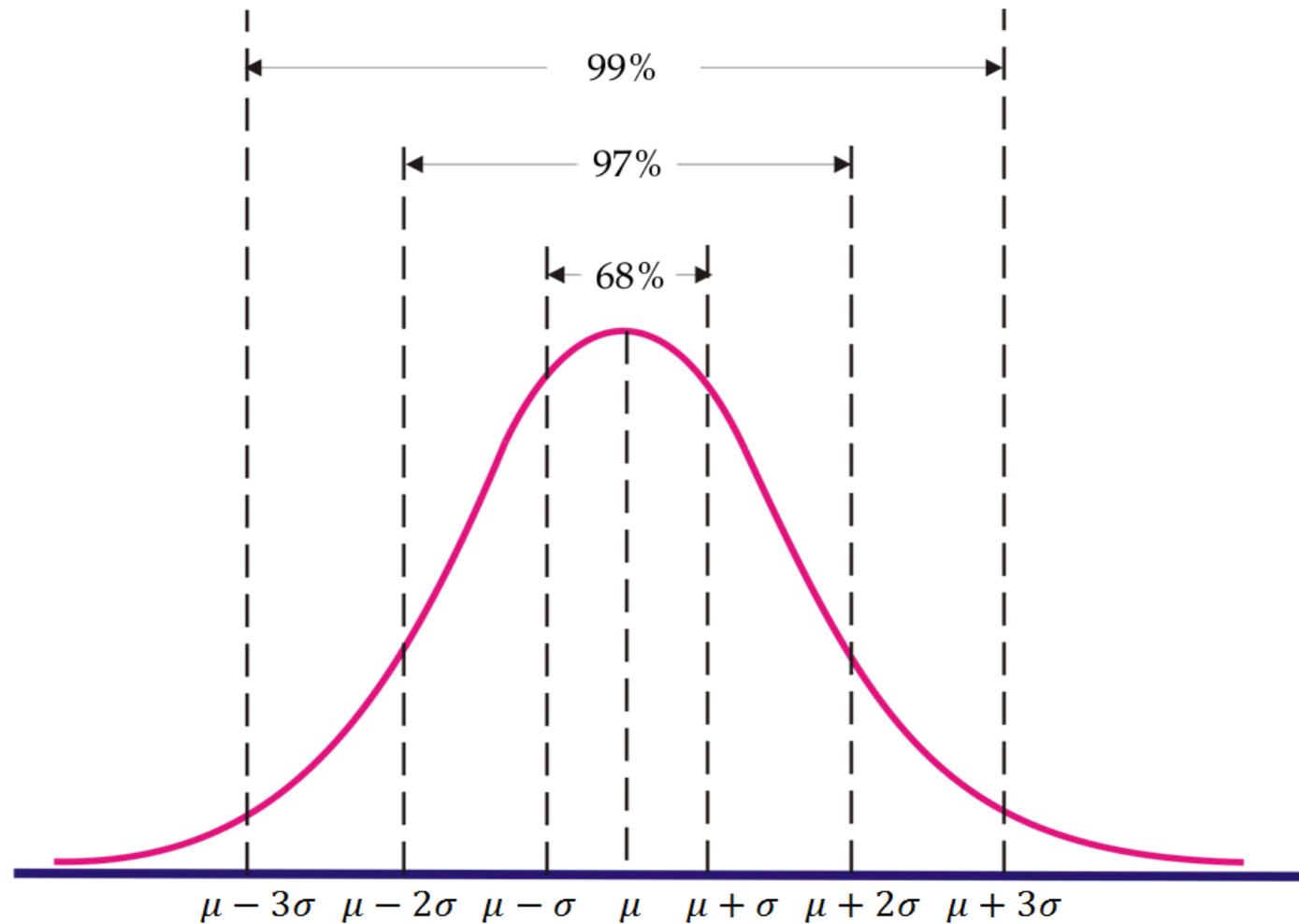
## Test de Anderson - Darling

Dado una muestra de datos previamente ordenada en forma ascendente  $x_1, x_2, \dots, x_n$  se busca contrastar estos valores con la función de distribución acumulada de la distribución normal, por medio de la siguiente expresión:

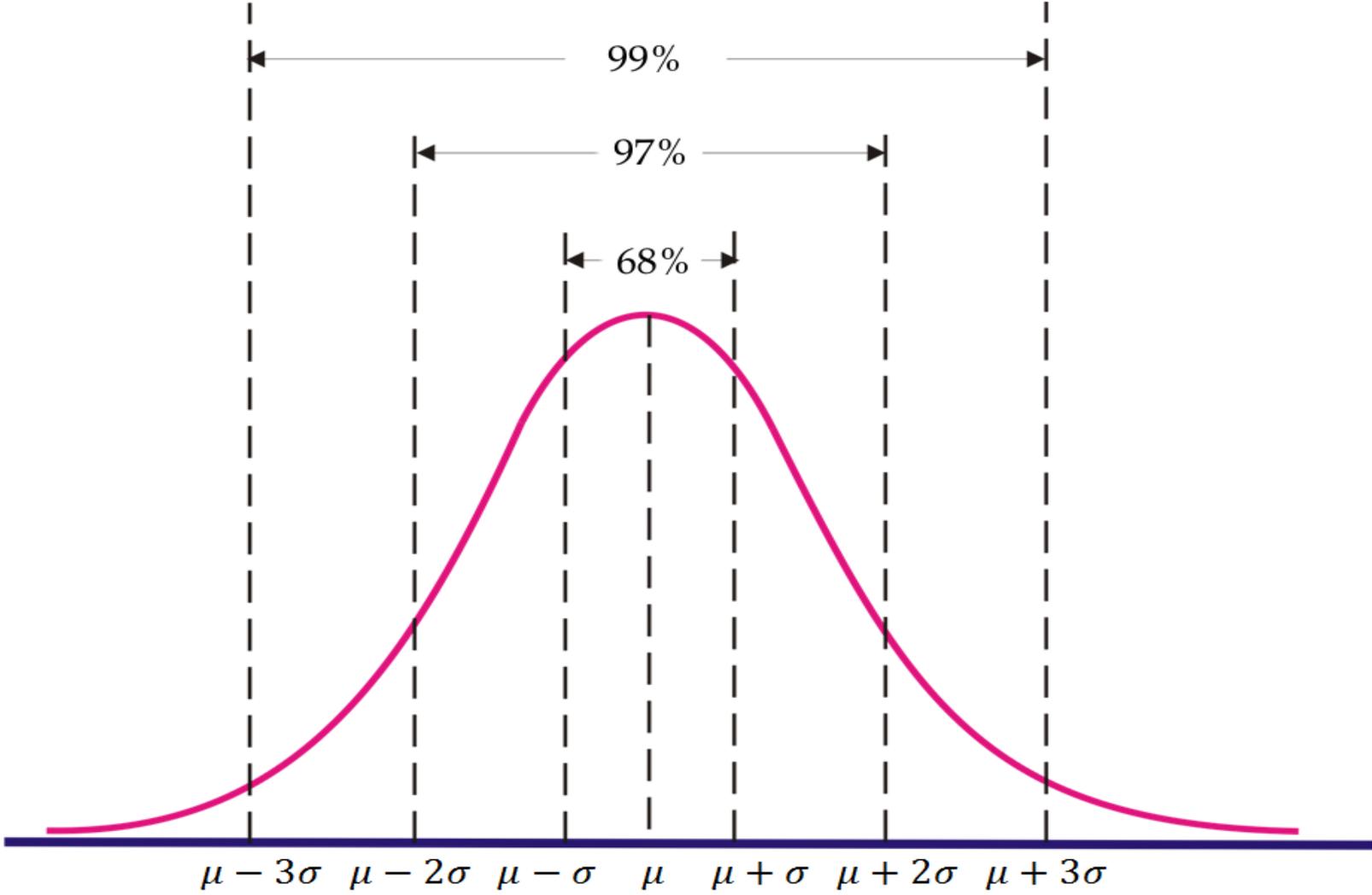
$$A^2 = -n - \sum_{j=1}^n \frac{2j-1}{n} \left[ \ln F(x_j) + \ln (1 - F(x_{n+1-j})) \right]$$

Este valor se compara con la distribución del estadístico de prueba (por ejemplo, la distribución normal) para determinar un valor crítico.

# Distribución Normal

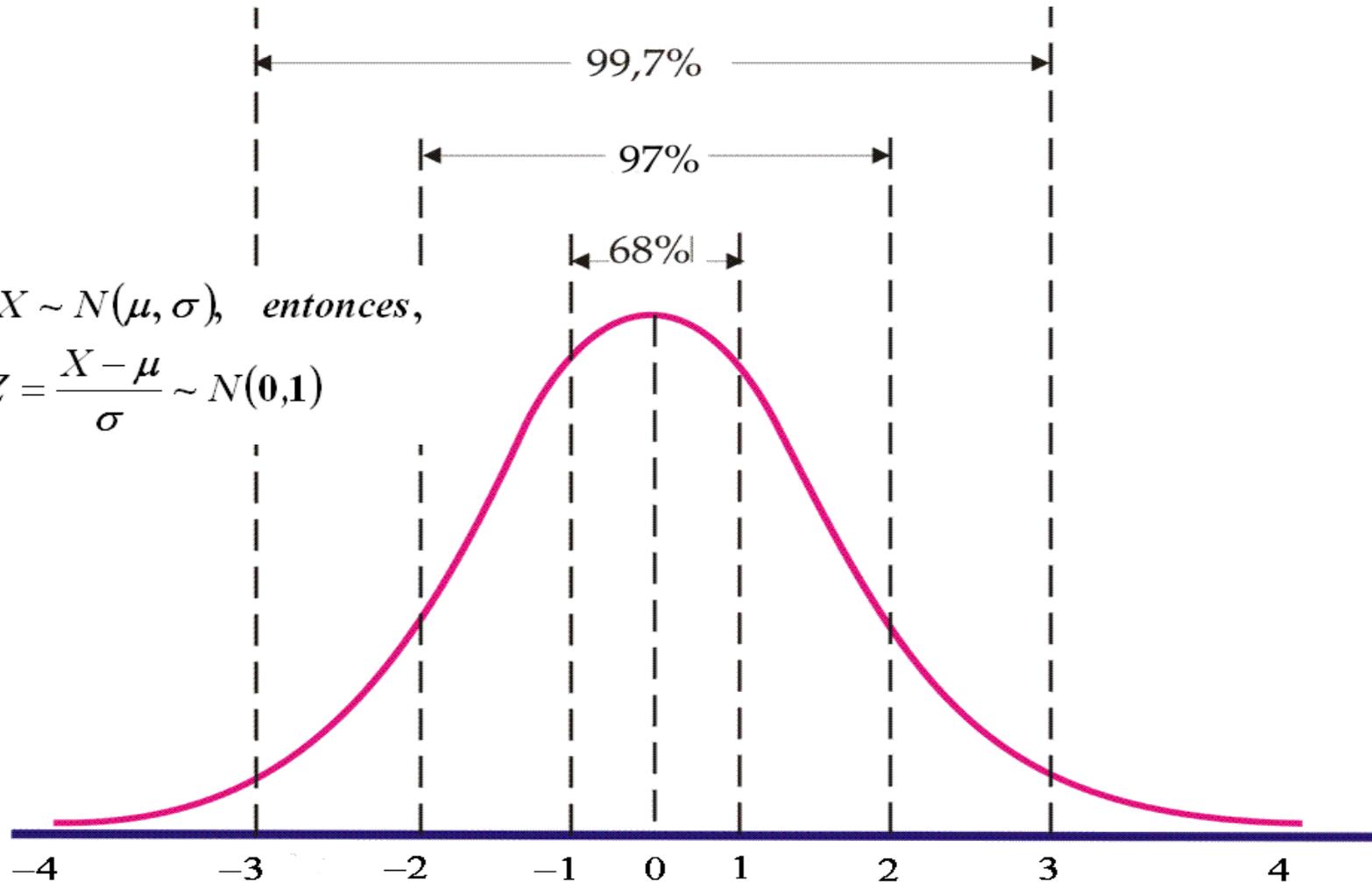


# Regla 68-97-99,7



# Regla 68-97-99,7

Si  $X \sim N(\mu, \sigma)$ , entonces,  
 $Z = \frac{X - \mu}{\sigma} \sim N(0,1)$

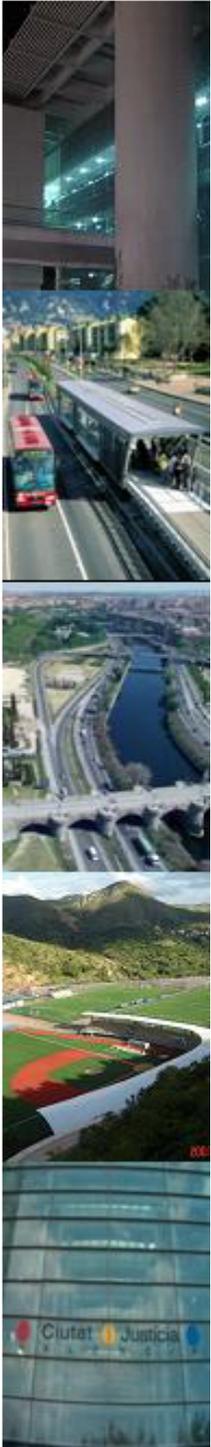




# Tabla de la Distribución Normal Estándar

normal	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	<b>0.8413</b>	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	<b>0.9505</b>	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2	<b>0.9773</b>	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	<b>0.9987</b>	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998





# Tabla de la Distribución Normal Estándar

normal	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-4	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.0001	0.0001	0.0001	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.0002	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.0003	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.0004	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.0006	0.00058	0.00056	0.00054	0.00052	0.0005
-3.1	<b>0.00097</b>	0.00094	0.0009	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.001
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.0024	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.0028	0.00272	0.00264
-2.6	0.00466	0.00453	0.0044	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.0057	0.00554	0.00539	0.00523	0.00508	0.00494	0.0048
-2.4	0.0082	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.0099	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.0139	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.0116	0.0113	0.01101
-2.1	0.01786	0.01743	0.017	0.01659	0.01618	0.01578	0.01539	0.015	0.01463	0.01426
-2	<b>0.02275</b>	0.02222	0.02169	0.02118	0.02068	0.02018	0.0197	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.0268	0.02619	0.02559	0.025	0.02442	0.02385	0.0233
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.0392	0.03836	0.03754	0.03673
-1.6	0.0548	0.0537	0.05262	0.05155	0.0505	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.0778	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.3	0.0968	0.0951	0.09342	0.09176	0.09012	0.08851	0.08692	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
-1.1	0.13567	0.1335	0.13136	0.12924	0.12714	0.12507	0.12302	0.121	0.119	0.11702
-1	<b>0.15866</b>	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.2327	0.22965	0.22663	0.22363	0.22065	0.2177	0.21476
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.2451
-0.5	0.30854	0.30503	0.30153	0.29806	0.2946	0.29116	0.28774	0.28434	0.28096	0.2776



# Tabla de la Distribución Normal Estándar

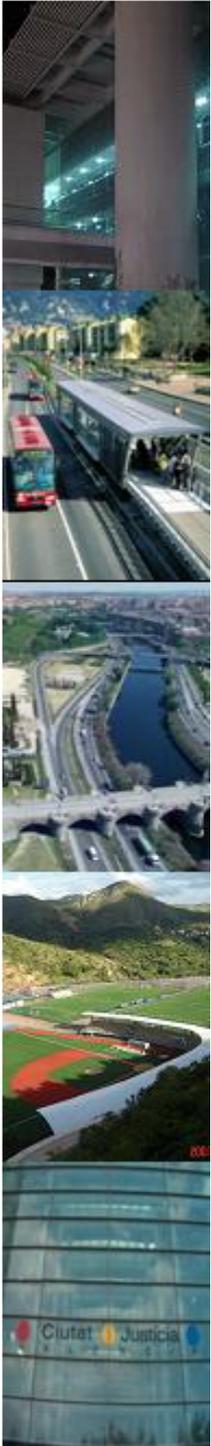
$$P[-z_0 < Z < z_0] = P[Z < z_0] - P[Z < -z_0]$$

$$P[-1 < Z < 1] = 0.8413 - 0.1586 = 0.6826$$

$$P[-2 < Z < 2] = 0.9775 - 0.0227 = 0.9545$$

$$P[-3 < Z < 3] = 0.9986 - 0.0135 = 0.9973$$

$$P[Z < 1.65] = 0.95$$



# Tabla de la Distribución Normal Estándar

$$Z = \frac{X - \mu}{\sigma} \approx N(0,1)$$

Con un **95%** de confianza puedo saber que:

$$Z = \frac{X - \mu}{\sigma} = 1.65 \rightarrow X = \mu + 1.65\sigma$$

$$CR = CB[\mu + 1.65\sigma],$$

$$\mu = 30\%, \sigma = 10\%, CB = 120 \quad (\text{Sobre costo})$$

$$CR = 120[0.30 + 1.65 \times 0.10] = 55.8$$

# Tabla de la Distribución Normal Estándar

$$Z = \frac{X - \mu}{\sigma} \approx N(0,1)$$

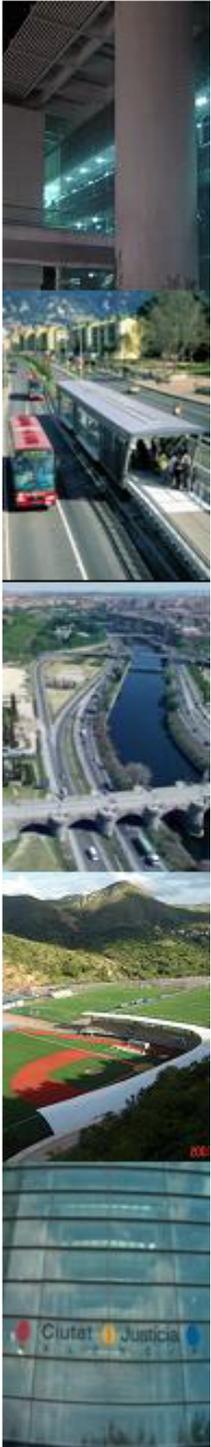
Con un 5% de confianza puedo saber que:

$$Z = \frac{X - \mu}{\sigma} = -1.65 \rightarrow X = \mu - 1.65\sigma$$

$$CR = CB[\mu - 1.65\sigma],$$

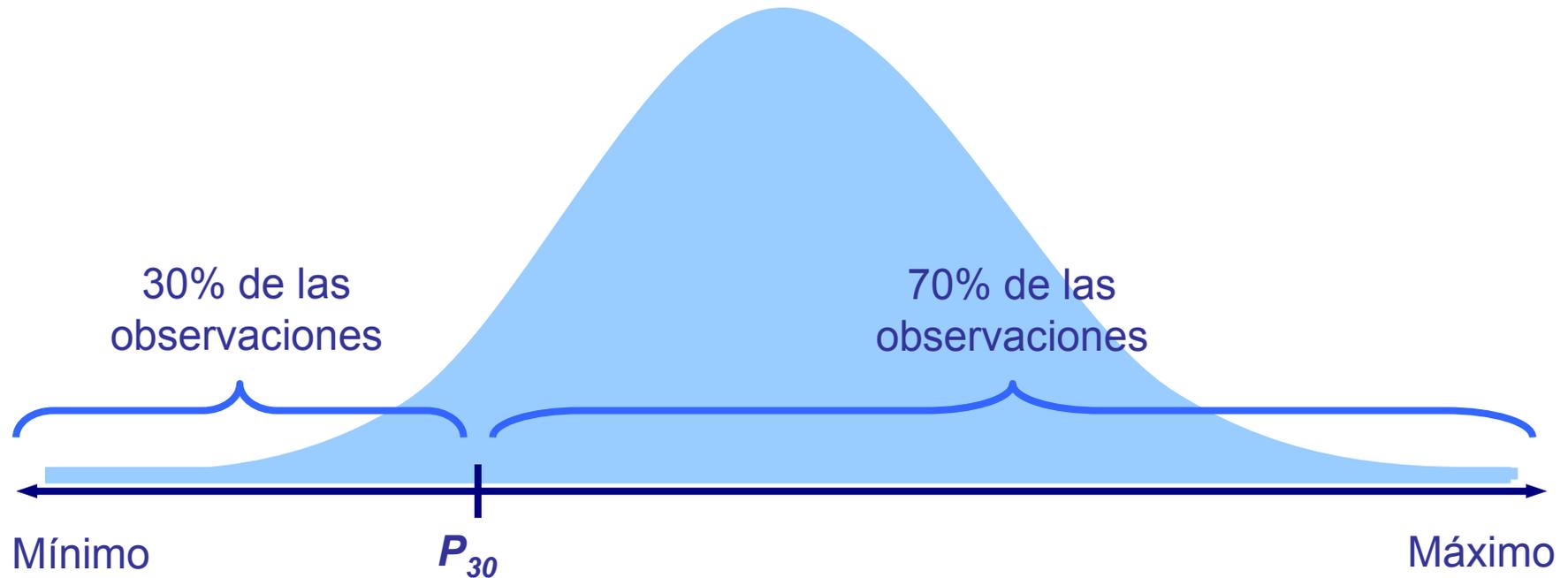
$$\mu = 30\%, \sigma = 10\%, CB = 120 \quad (\text{Sobre costo})$$

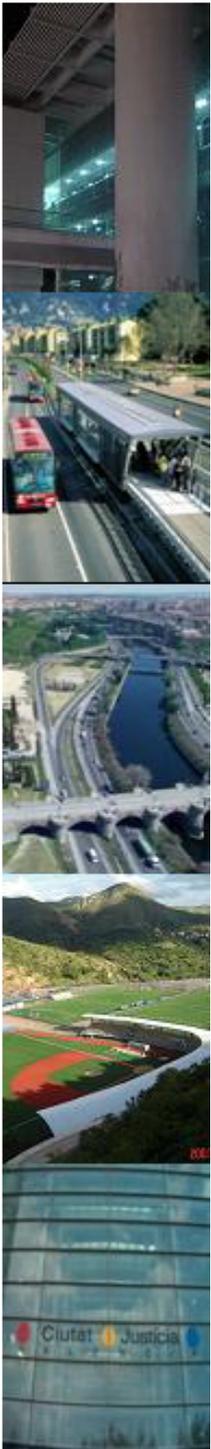
$$CR = 120[0.30 - 1.65 \times 0.10] = 16.2$$



## Percentil

El percentil es una medida de posición central que divide a la distribución de probabilidad en dos partes. Por ejemplo, el percentil 30, ***P30***, es el valor que indica que por debajo de él, se encuentra el 30% de las observaciones y por encima de él se encuentra el 70% de las observaciones.





# Cálculo del Percentil – Datos Ordenados

$$L_p = (n + 1) \frac{P}{100}$$

$L_p$  : Ubicación del percentil

$n$  : Número de observaciones

$P$  : Percentil deseado

Consideremos las siguientes muestras de sobrecostos de un muestra de proyectos (en 100's)

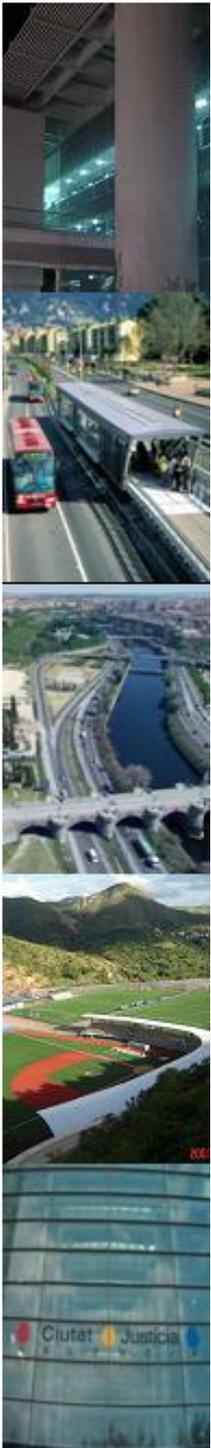
3	10	19	27	34	38	48	56	67	74
4	12	20	29	34	39	48	59	67	74
7	14	21	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

$$L_{60} = (50 + 1) \frac{60}{100} = 30,6$$

El valor resultante de 30,6 dice que el percentil 60 esta ubicado al 60% del trayecto comprendido entre la treintava observación, que es 47 y la treintaiunava observación que es 48, es decir:

$$P_{60} = 47 + 0,6(48 - 47) = 47,6$$

Por lo tanto, el 60% de las observaciones está por debajo de 47,6 y el 40% restante por encima de 47,6



## Cálculo del Percentil – Datos Agrupados

$$P_k = y'_{j-1} + \left( \frac{\frac{kn}{100} - F_{j-1}}{f_j} \right) \times A$$

$k$  : Percentil deseado

$y'_{j-1}$  : Límite inferior exacto de la clase que contiene el percentil deseado.

$A$  : Ancho del intervalo

$n$  : Frecuencia total

$F_{j-1}$  : Frecuencia acumulada de la clase anterior a la que contiene al percentil deseado

$f_j$  : Frecuencia absoluta de la clase que contiene el percentil deseado

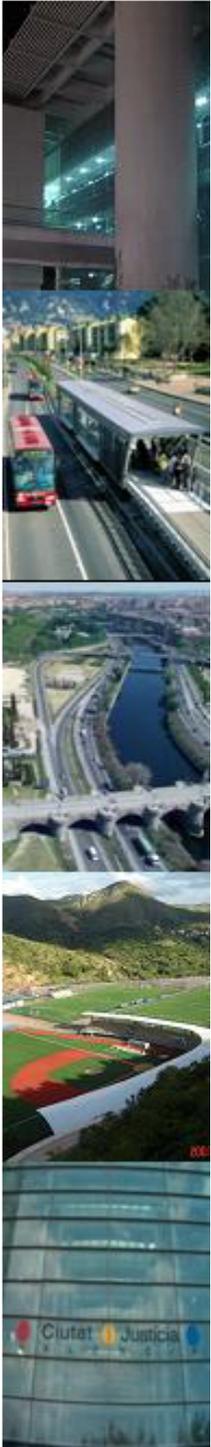
$y_{i-1} - y_i$	$f_j$	$F_j$
30,1 - 38	9	9
38,1 - 46	16	25
46,1 - 54	31	56
54,1 - 62	42	98
62,1 - 70	23	121
70,1 - 78	15	136
78,1 - 86	9	145
86,1 - 94	5	150
$\Sigma$	150	-

←  $F_{j-1}$

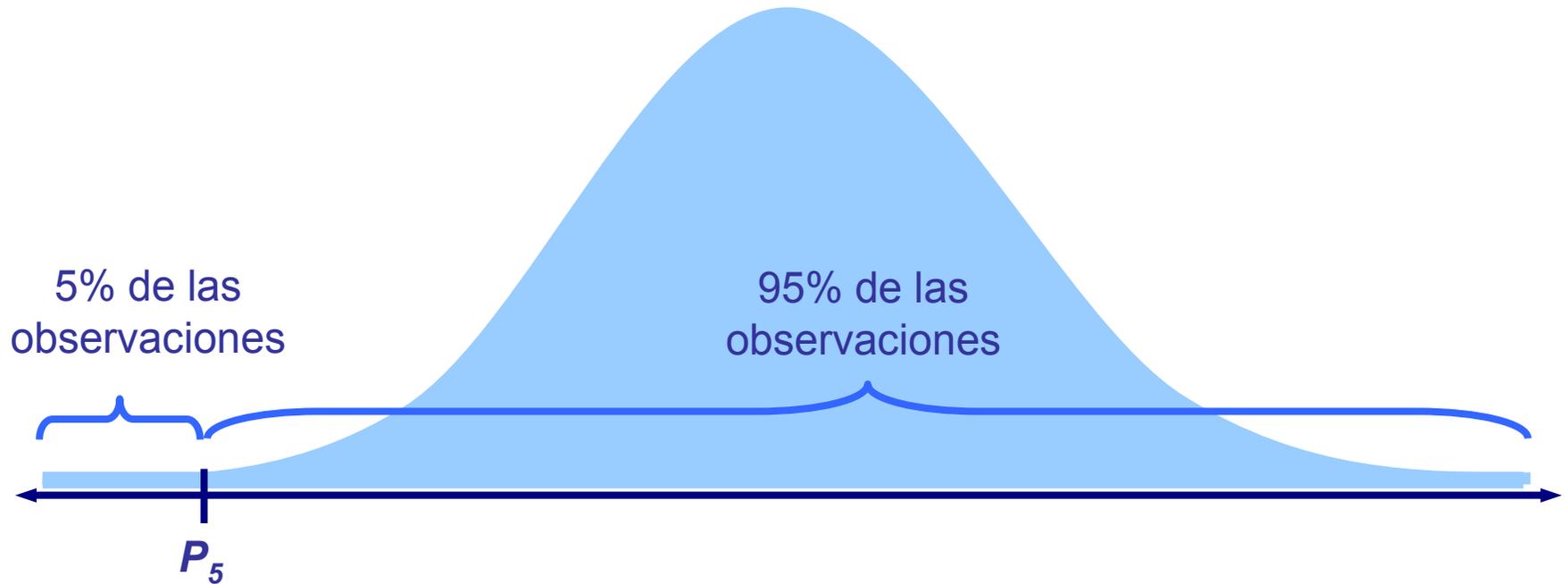
←  $F_j$

$$P_{75} = 62,1 + \left( \frac{\frac{75 \times 150}{100} - 98}{23} \right) \times (70 - 62,1) = 67,1$$

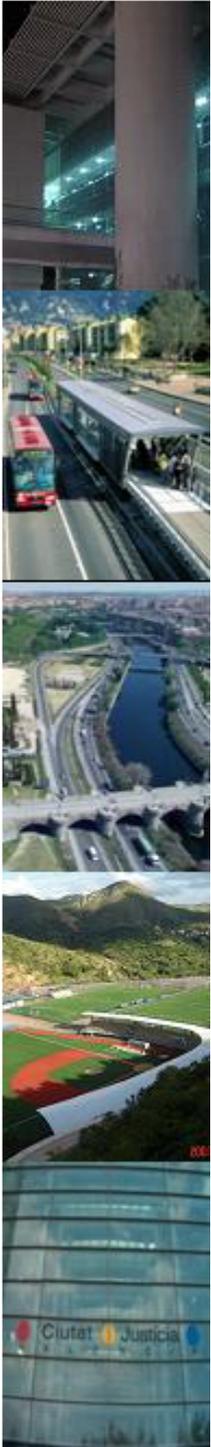
Por lo tanto, el 75% de las observaciones está por debajo de 67,1 y el 25% restante por encima de 67,1



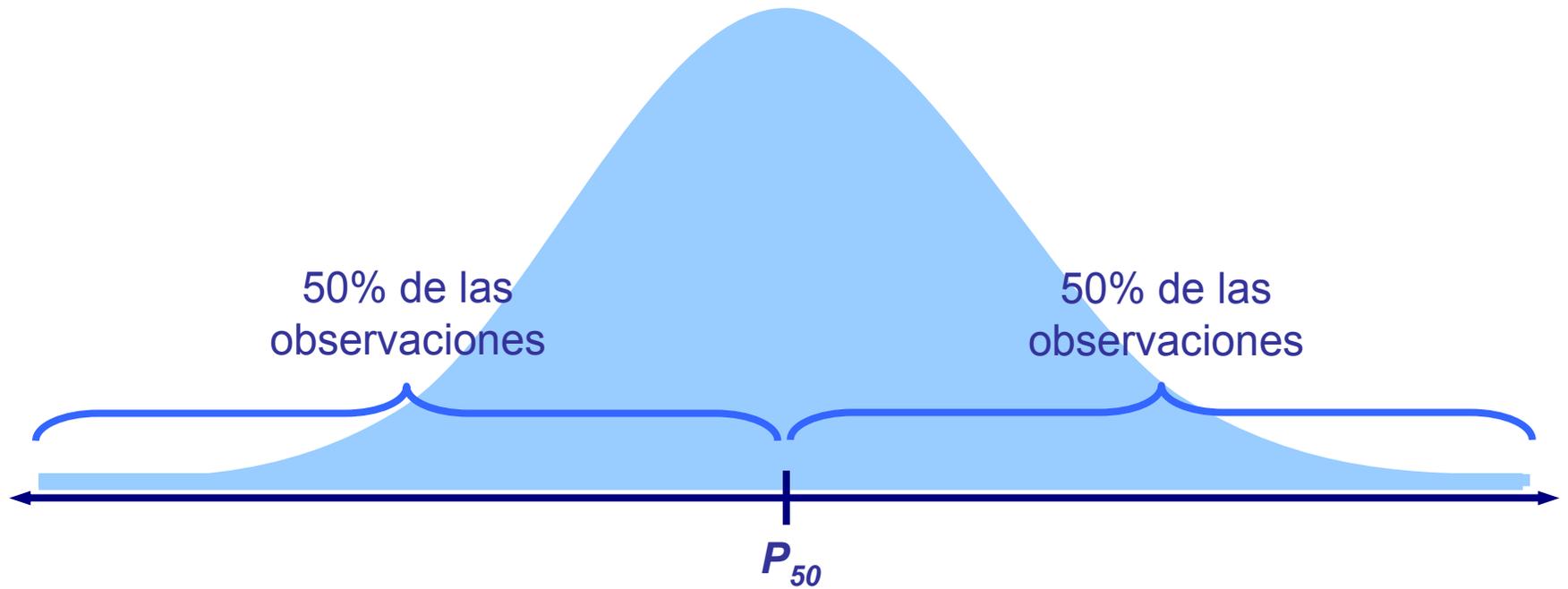
## Percentil 5



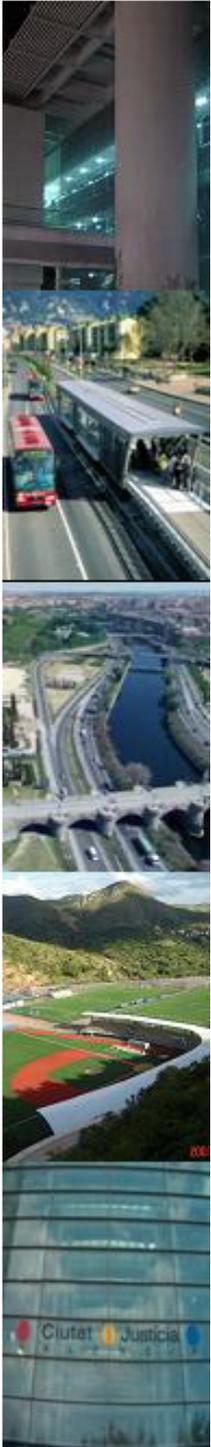
$$P_5 = y_{j-1} + \left( \frac{\frac{5 \times n}{100} - F_{j-1}}{f_j} \right) \times A$$



# Percentil 50



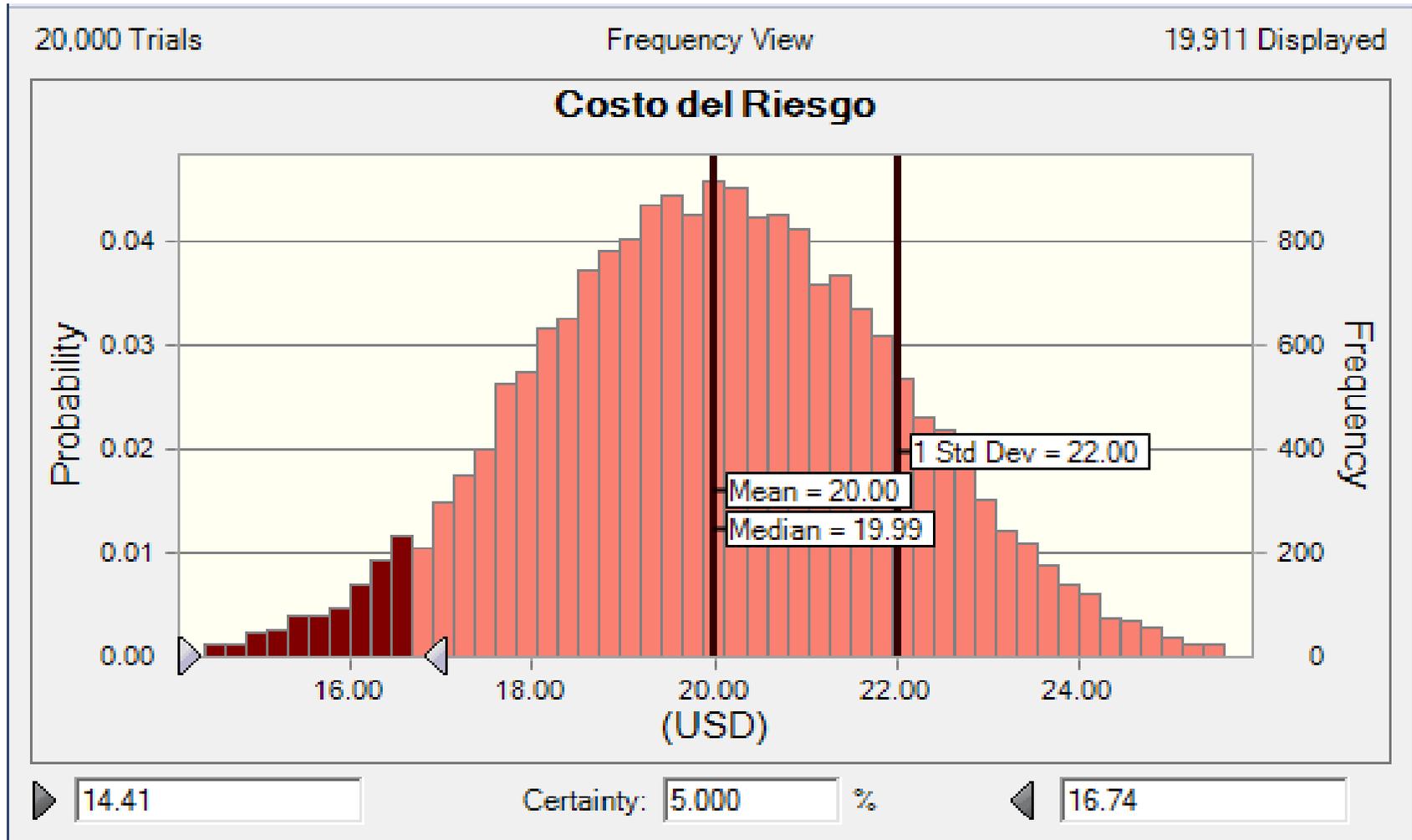
$$P_{50} = y_{j-1} + \left( \frac{\frac{50 \times n}{100} - F_{j-1}}{f_j} \right) \times A$$

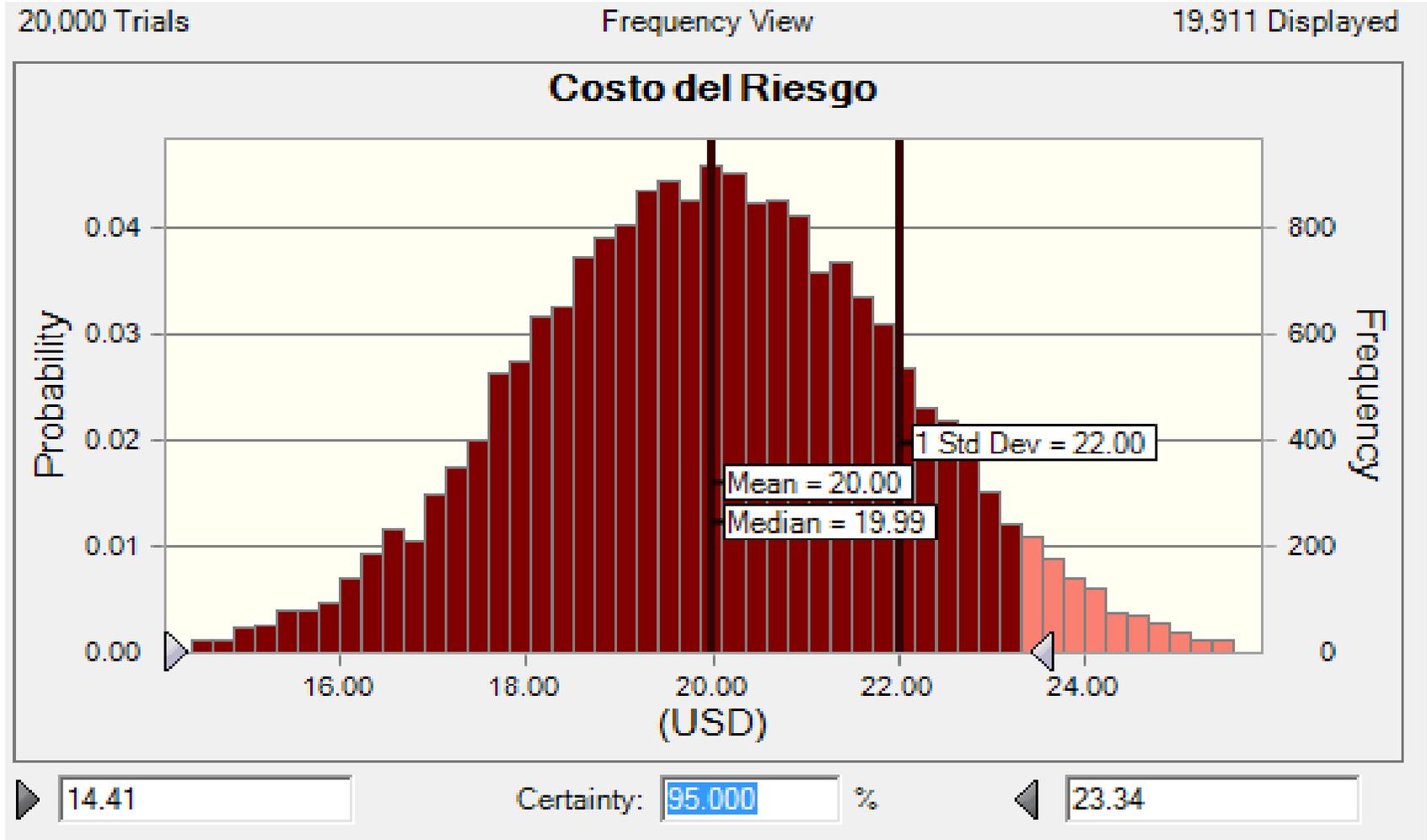
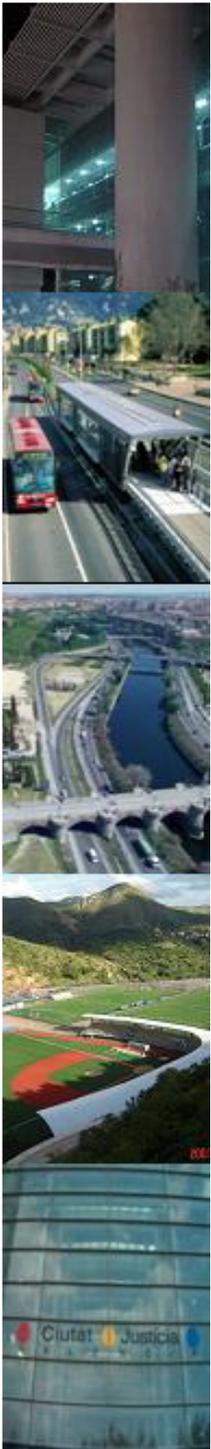


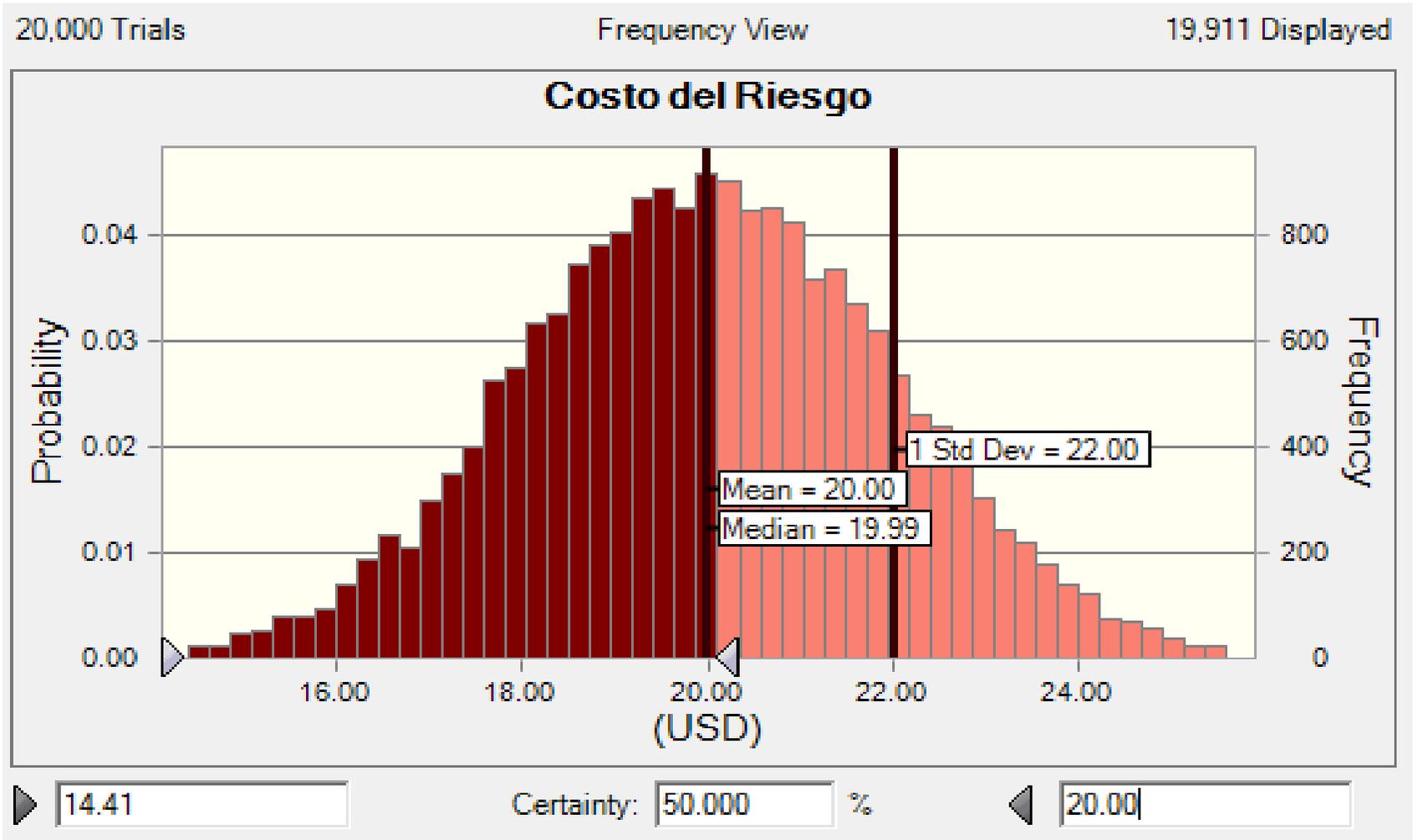
# Percentil 95

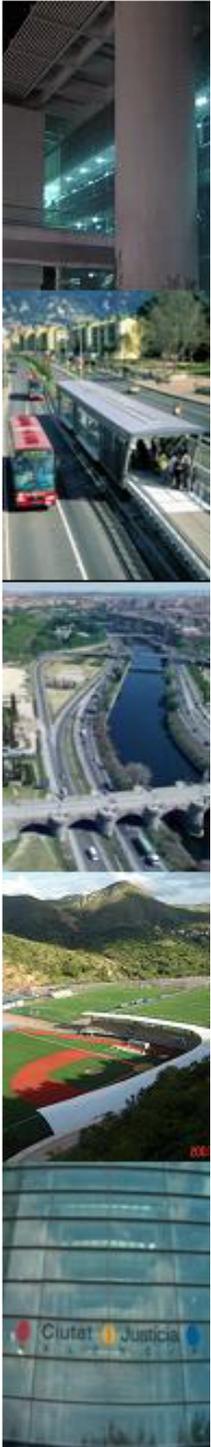


$$P_{95} = y_{j-1} + \left( \frac{\frac{95 \times n}{100} - F_{j-1}}{f_j} \right) \times A$$





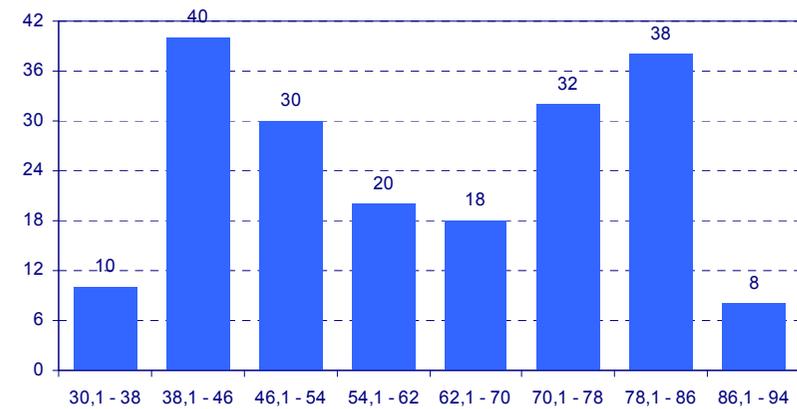
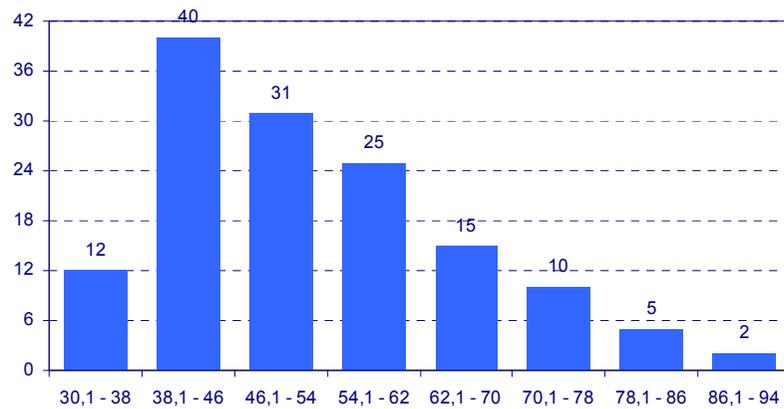


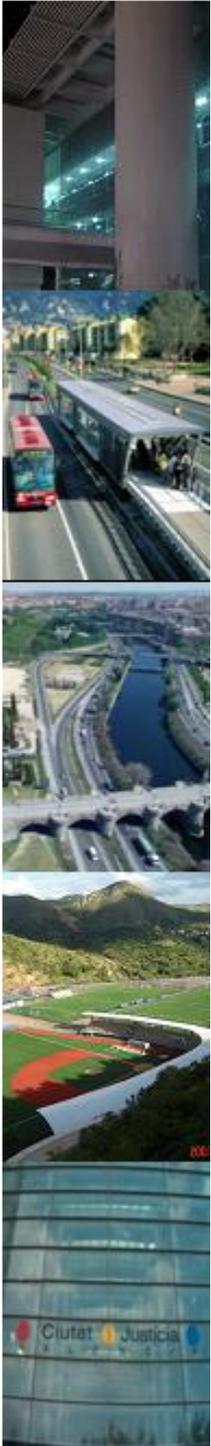


# Definición del Método Bootstrap

El método Bootstrap fue propuesto por Efron (1977), estadístico de la universidad de Stanford. Se utiliza cuando no se tiene ninguna idea acerca de la posible distribución de los datos o cuando estos no siguen ninguna distribución conocida. De hecho, esta técnica permite obtener más información de los datos que cualquier otro método estadístico previamente desarrollado, resolviendo el problema clave en estadística, que es inferir la “verdad” de una muestra que se sabe es incompleta. **Tomando en consideración el Teorema Central del Límite, las medias de cada muestra generada convergen a una distribución de probabilidad normal.**

Siempre que se cuente con una muestra pequeña ( $N < 20$ ) la cual no permite inferir el tipo de distribución de probabilidad que está asociada a la muestra, se procede a emplear una técnica de remuestreo, la cual trata a la muestra como si fuera una población, esta técnica es conocida como el Método Bootstrap (Estimación Autosuficiente o remuestreo).

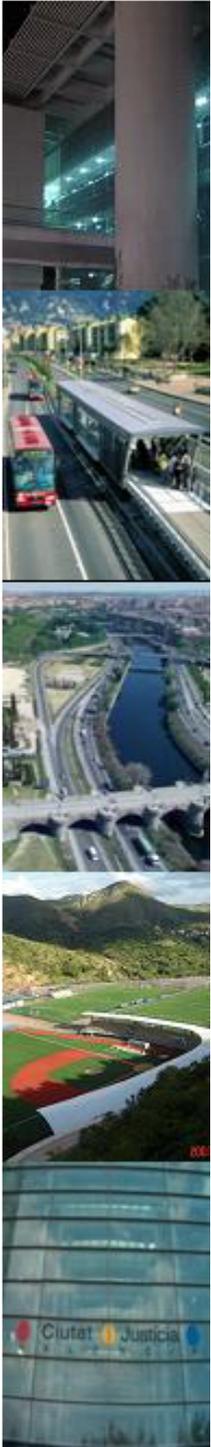




## Proceso del Método Bootstrap

1. Sea  $X_1, X_2, \dots, X_n$  la muestra inicial.
2. Por medio de un generador de números aleatorios obtenemos una submuestra  $X_1^*, X_2^*, \dots, X_n^*$  independiente y con reemplazo de la muestra original, a esta nueva muestra se le conoce como muestra Bootstrap.
3. Se procede a calcular el estadístico de interés para la muestra Bootstrap, por ejemplo la media  $\mu_1^*$ .
4. Se repite los pasos 1 y 2 una cantidad muy grande de veces,  $N$ , de donde se obtiene los estadísticos  $\mu_1^*, \mu_2^*, \dots, \mu_N^*$ .
5. Con esta nueva muestra se procede a realizar todo proceso descriptivo e inferencial.

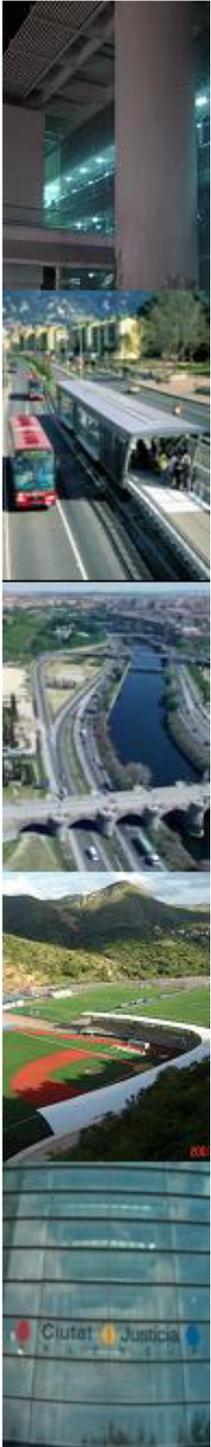
Fuente: Adaptado de DiCiccio y Tibshirani (1987).



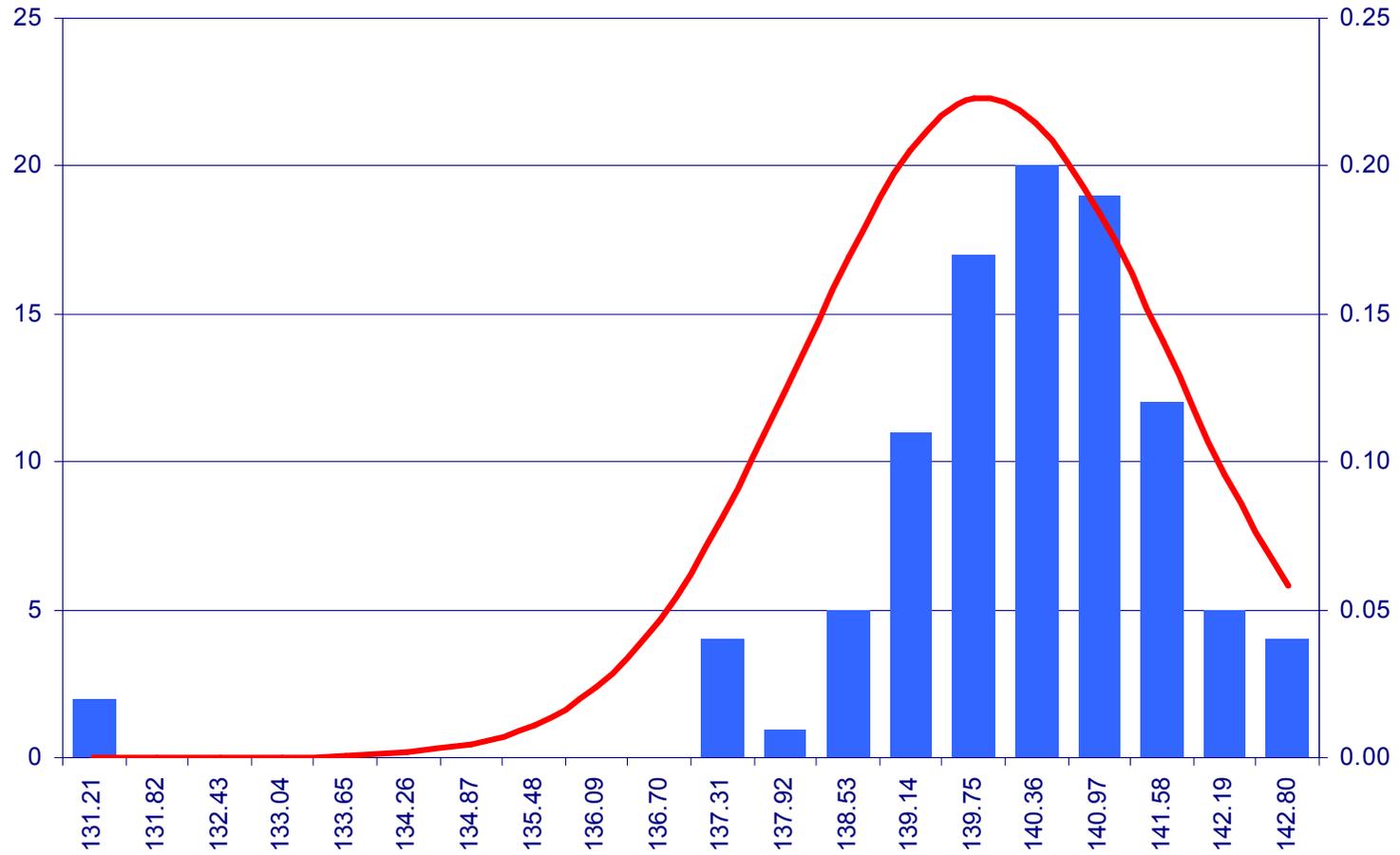
## Ejemplo del Método Bootstrap

Para ilustrar el uso de este método se mostrarán cien muestras Bootstrap generadas a partir de 20 datos.

Muestra	Dato 1	Dato 2	Dato 3	Dato 4	Dato 5	...	Dato 20	$\bar{X}$	$\sigma$
Dato Original	129	132	147	149	153	...	151	142,0	60,1
Bootstrap 1	149	147	147	132	132	...	151	140,7	59,1
Bootstrap 2	149	129	149	153	149	...	129	139,8	58,7
Bootstrap 3	129	147	153	149	132	...	132	140,1	62,1
Bootstrap 4	153	153	129	153	132	...	132	140,4	60,5
Bootstrap 5	147	149	129	147	147	...	151	140,5	60,5
Bootstrap 6	132	149	149	132	132	...	132	137,2	59,3
Bootstrap 7	153	153	147	153	153	...	153	140,5	58,9
Bootstrap 8	132	149	153	132	147	...	151	140,4	61,6
Bootstrap 9	129	153	149	129	147	...	147	140,5	58,1
Bootstrap 10	129	129	149	132	129	...	129	140,3	59,1
...	...	...	...	...	...	...	...	...	...
Bootstrap 1000	129	129	147	132	132	...	129	141,4	56,3



# Ejemplo del Método Bootstrap



Media = 139.86

Desviación Estándar = 1.79

