FUNDAMENTOS BÁSICOS DE ESTADÍSTICA

David de Bustamante, MFIA

Director de Mercado

MEFF

INSTITUTO BME

Fundamentos básicos de estadística

1. Variables aleatorias y distribuciones de frecuencia

- a. Variable aleatoria
- b. Histogramas de Frecuencia.
- c. Medidas de tendencia central: Media, moda y mediana
- d. Medias de dispersión: varianza y volatilidad
- e. Desigualdad de Tchebychev
- f. Otros momentos: Asimetría y curtosis.

2. Vectores aleatorios y distribuciones de frecuencia conjuntas. Medidas estadísticas

- a. Covarianza
- b. Matriz de Varianzas y covarianzas
- c. Algebra matricial
- d. Coeficiente de correlación

Fundamentos básicos de estadística

3. Modelos de Regresión simple

- a. Estimación de parámetros
- b. MCO
- c. Fiabilidad de estimaciones
- d. Intervalo de confianza
- e. Contraste de Hipótesis
- f. Coeficiente de Determinación
- g. Función: estimación lineal

Conceptos básicos de estadística

Una **variable aleatoria** es una variable que puede tomar una serie de valores, cada uno de ellos con una cierta probabilidad. La variables de los mercados financieros se comportan de manera aleatoria ya que sus valores futuros son impredecibles. Pueden ser:

- **Discretas**: El número de valores que puede tomar la variable es finito, por ejemplo el lanzamiento de una moneda o dado.

- **Continuas**: El número de valores que puede tomar la variable es infinito (no conocido a priori), por ejemplo el valor del IBEX 35 dentro de una año.

Histograma de frecuencias

Cuando se trata de analizar la dispersión que presentan unos datos, la representación gráfica más adecuada es el histograma. Para realizar un histograma se marcan una serie de intervalos sobre un eje horizontal, y sobre cada intervalo se coloca un rectángulo de altura proporcional al número de observaciones (frecuencia absoluta) que caen dentro de dicho intervalo. De esta manera el histograma de frecuencias resulta muy útil para representar gráficamente la distribución de frecuencias

Si se pretende comparar varios histogramas construidos con distinto número de datos, es preferible que las alturas de los rectángulos sean proporcionales al porcentaje de observaciones en cada intervalo o al tanto uno por uno (frecuencia relativa). Utilizando la frecuencia relativa en el eje de ordenadas también se facilita la comparación entre el histograma obtenido y un determinado modelo teórico representado por una función densidad de probabilidad. En este caso se considera que la frecuencia relativa es proporcional al área definida por cada columna. Puede interpretarse la función densidad de probabilidad como la representación del histograma cuando el número de observaciones tiende a infinito y la anchura de los rectángulos tiende a cero.

Histograma de frecuencias

TEF	251			
16,35				
16,08	-1,67%	MAX	6,76%	
16,20	0,74%	MIN	-7,12%	
16,26	0,37%	Rango	13,88%	
16,00	-1,61%	Clases	16	
15,97	-0,19%	Longitud	0,87%	
16,03	0,38%			
16,05	0,12%			
16,25	1,24%			
16,26	0,06%			
16,10	-0,99%		-7,12%	=+FRECUENCIA(C3:C253;E12:E28)
16,16	0,37%		-6,25%	0,00
16,05	-0,68%		-5,39%	1,00
16,22	1,05%		-4,52%	1,00
16,10	-0,74%		-3,65%	1,00
16,27	1,05%		-2,79%	1,00
16,25	-0,12%		-1,92%	17,00
16,25	0,00%		-1,05%	30,00
16,38	0,80%		-0,18%	59,00
16,35	-0,18%		0,68%	65,00
16,41	0,37%		1,55%	39,00
16,66	1,51%		2,42%	24,00
16,55	-0,66%		3,29%	5,00
16,47	-0,48%		4,15%	4,00
16,49	0,12%		5,02%	2,00
16,60	0,66%		5,89%	0,00
16,51	-0,54%		6,77%	1,00
16,63	0,72%			

El rango nos informa de la extensión de la variabilidad de los datos:

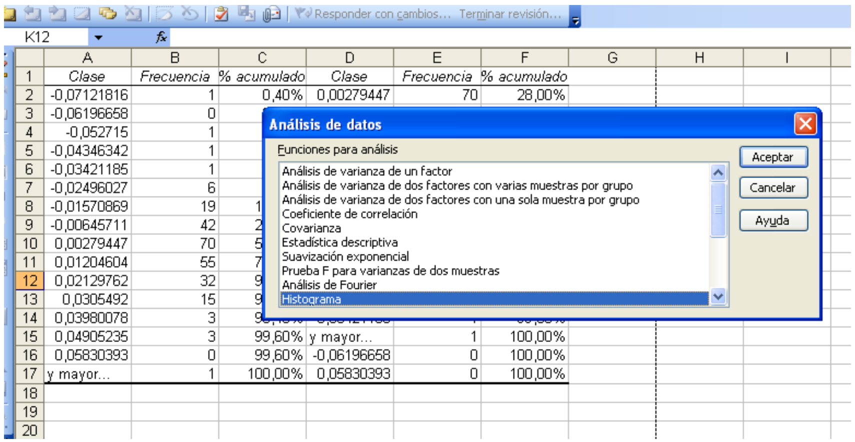
Rango=Max (Xi) - Min(Xi)

Número de clases: Raiz(N) (se redondea)

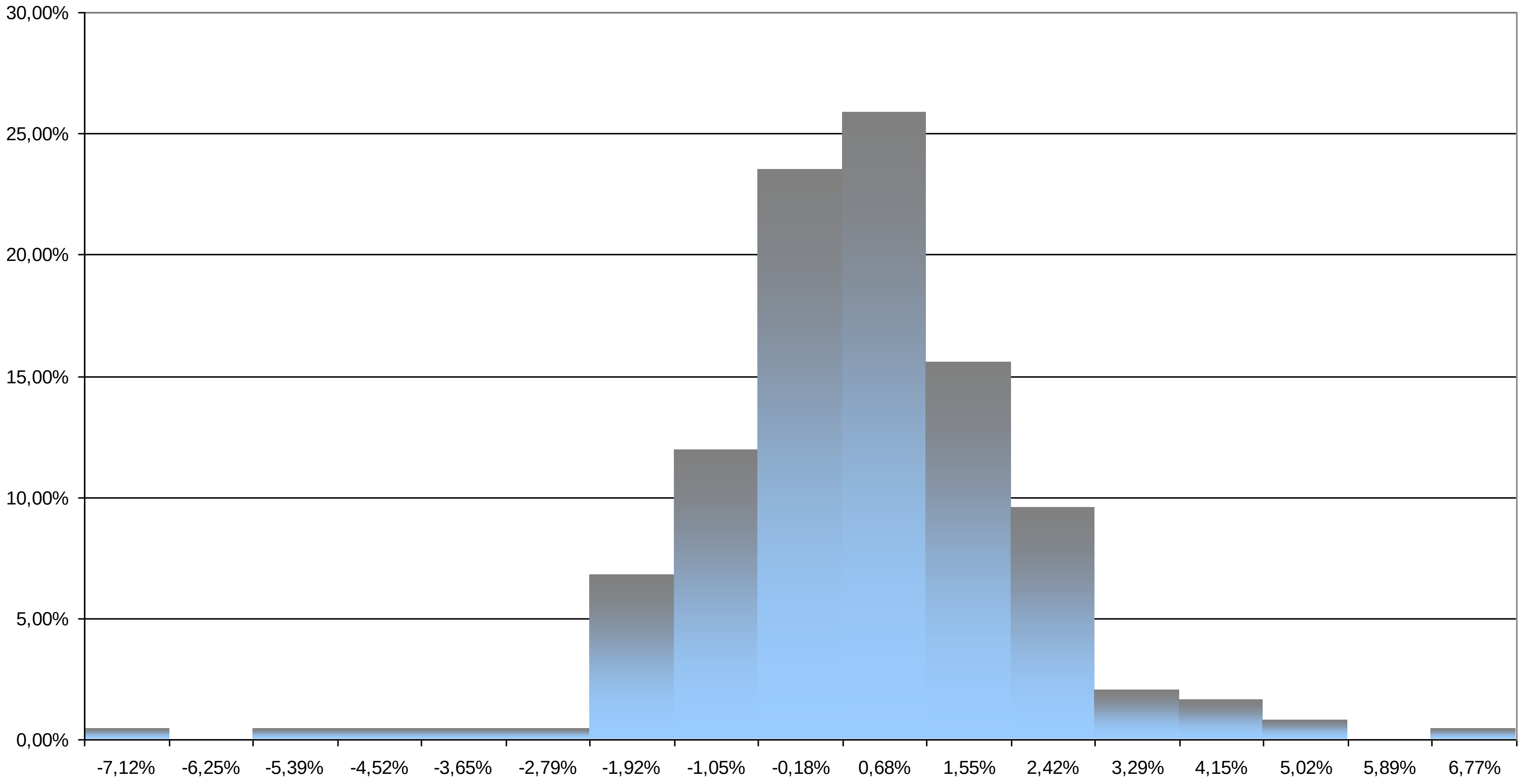
Longitud de las clases: Rango / Número de clases

Si el número de clases por la longitud supera el rango se ajusta dividiendo la diferencia entre dos y ampliando la primera y última clase.

En Excel podemos utilizar la función “frecuencia” o por medio del “análisis de datos”



Histograma de frecuencias



Medidas de tendencia Central y No Central

Las medidas de una variable aleatoria son:

1. **Centrales**:

a. Media: promedio de los valores posibles de la v.a.:

Para v.a. continuas: $$E(x) = \mu = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$$

Para v.a. discretas: $$E(x) = \mu = \sum_i x_i P(x) \text{ si son equiprobab les entonces } \mu = \frac{\sum_i x_i}{n}$$

b. Moda: Valor que se repite con mayor frecuencia.

c. Mediana: Valor central. Existen el mismo número de valores a la derecha que a la izquierda.

2. **No centrales**:

a. Cuartiles: Dividen la distribución en cuatro partes iguales.

b. Percentiles: Dividen la distribución en x partes de 100.

Medidas de Dispersión

1a. **Varianza**: La varianza es la medida de cómo están de concentrados o dispersos los valores alrededor de media. Su valor es la media del cuadrado de las desviaciones respecto a su media, se eleva al cuadrado para compensar valores positivos y negativos. Es el momento de orden dos (por ser al cuadrado) mientras que la media es el momento de orden uno.

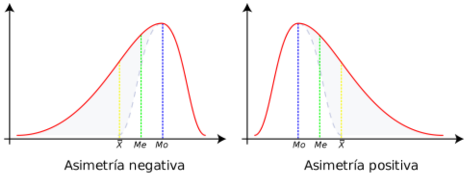
Para v.a discretas: $$Var(x) = \sigma^2 = \sum_i (x_i - E(x))^2 \cdot P(x) \text{ ó } \frac{\sum_i (x_i - E(x))^2}{n}$$

Para v.a. continuas: $$Var(x) = \sigma^2 = \int_{-\infty}^{\infty} (x - E(x))^2 \cdot f(x) \cdot dx$$

1b. **Desviación típica:** es la raíz de la varianza.

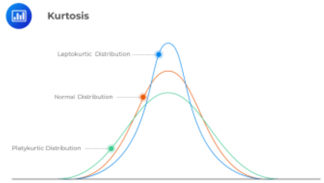
2. **Asimetría**: Momento de tercer orden. Indica la similitud de la distribución a la izquierda y a la derecha de la media.

$$\text{Coeficiente de Asimetría} = \frac{\sum (X_i - \bar{x})^3}{n\sigma^3}$$



3. **Curtosis**: Momento de cuarto orden. Indica la concentración de la v.a. en valores extremos alejados de la media. Como la curtosis de una distribución normal (0,1) es 3 es normal representar el exceso de curtosis (K-3).

$$\text{Coeficiente de Apuntamiento} = \frac{\sum (X_i - \bar{x})^4}{n\sigma^4}$$



Medidas de Tendencia y Dispersión
Información que aportan

La información que nos aportan en conjunto la media y la desviación típica viene dada por la desigualdad de **Tchebychev**.

Viene a decirnos el porcentaje de datos que hay entre la media y ”+/-k” veces la desviación típica.

$$\left(1 - \frac{1}{k^2}\right)$$

Formalmente la **desigualdad de Tchebychev**:

$$P(E(x) - k\sigma \leq x \leq E(x) + k\sigma) \geq 1 - \frac{1}{k^2}$$

Medidas estadísticas para Vectores Aleatorios

Cuando tenemos dos variables aleatorias y necesitamos medir la relación entre ellas utilizamos la **Covarianza**:

$$Cov(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i-\bar{x})(y_j-\bar{y})\cdot P(x,y)$$

La covarianza mide cómo fluctúan cada una de las variables entorno a su media. Al igual que la varianza, está expresada en puntos básicos al cuadrado y depende de las unidades de medida, por ejemplo si estuviéramos estudiando altura y edad seria altura x edad. Su interpretación es muy difícil, por ello se calcula el **coeficiente de correlación**:

$$\rho=\frac{Cov(x,y)}{\sigma_x\cdot\sigma_y}$$

- Es adimensional y toma valores entre +1 y -1. +1 indica una relación directa perfecta entre las variables y -1 una relación indirecta perfecta.
- Cuando el valor es 0 indica que no existe relación lineal entre las variables.

Medidas estadísticas para Vectores Aleatorios

Varianza y Desviación típica de una cartera:

Cartera con 2 activos :

$$\sigma_c^2 = w_a^2 \cdot \sigma_a^2 + w_b^2 \cdot \sigma_b^2 + 2 w_a w_b Cov_{a,b}$$

Cartera con n activos :

$$\sigma_c^2 = \begin{pmatrix} w_1 & w_2 & \dots & w_n \end{pmatrix} \cdot \begin{pmatrix} \sigma_1^2 & Cov_{1,2} & \dots & Cov_{1,n} \\ Cov_{2,1} & \sigma_2^2 & \dots & Cov_{2,n} \\ \dots & \dots & \dots. & \dots \\ Cov_{n,1} & Cov_{n,2} & & \sigma_n^2 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix}$$

(un apunte de algebra matricial)

Una matriz de orden m x n es un conjunto de elementos dispuestos en m filas y n columnas. Existen matrices cuadradas (m x m), rectangulares (m x n) y vectores fila o columna).

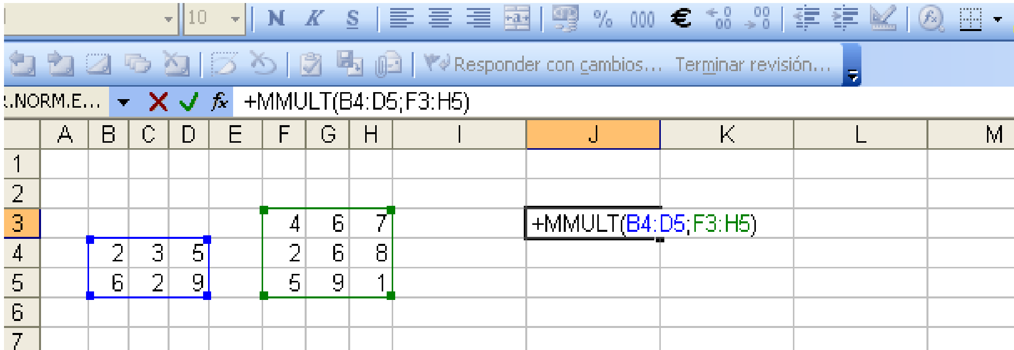
Suma: Se puede definir la suma de dos matrices siempre y cuando las dos matrices tengan el mismo orden, es decir, que ambas tengan el mismo número de filas y el mismo número de columnas, no quiere esto decir que sean cuadradas, sino que $m_a=m_b$ y $n_a=n_b$.

$$\begin{pmatrix} 2 & 4 \\ 5 & 6 \end{pmatrix}_{(2x2)} + \begin{pmatrix} 3 & 5 \\ 6 & 7 \end{pmatrix}_{(2x2)} = \begin{pmatrix} 5 & 9 \\ 11 & 13 \end{pmatrix}_{(2x2)}$$

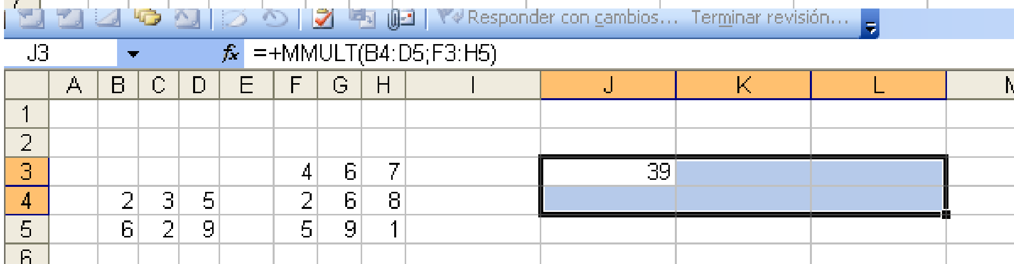
Producto: Dos matrices, aunque no sean cuadradas se pueden multiplicar cuando la primera matriz tenga un numero de columnas n, que coincida con el número de filas de la segunda matriz m.

$$\begin{pmatrix} 1 & 3 & 5 \\ 4 & 5 & 7 \end{pmatrix}_{(2x3)} \times \begin{pmatrix} 5 & 3 & 2 \\ 5 & 8 & 3 \\ 4 & 2 & 1 \end{pmatrix}_{(3x3)} = \begin{pmatrix} 40 & 37 & 16 \\ 73 & 66 & 30 \end{pmatrix}_{(2x3)}$$

Multiplicación de Matrices

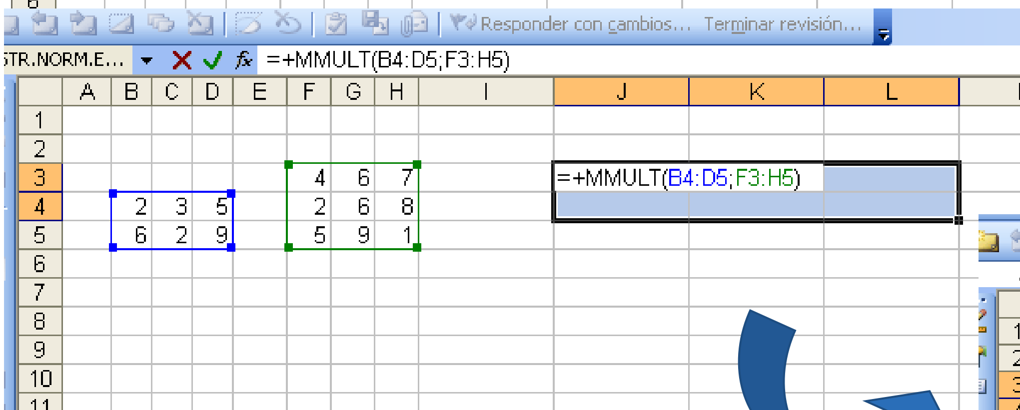


Utilizar la función MMULT y seleccionar las dos matrices a multiplicar. Intro

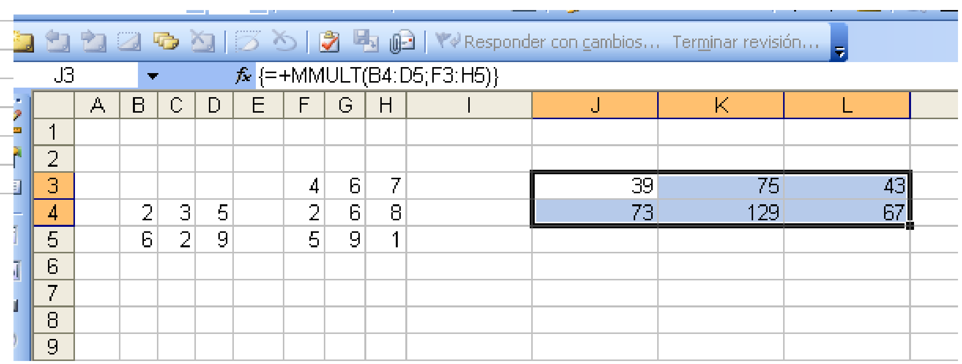


sombrear el numero de filas y columnas de la matriz resultante.

Pulsar F2 para editar la fórmula



Presionar simultáneamente las teclas CONTROL + SHIFT+ INTRO



Regresión Lineal

Uno de los principales objetivos de la regresión es estimar una variable (variable dependiente) en función de otra (variable independiente). Si estimamos y en función de x por medio de una ecuación, ésta se denomina ecuación de regresión.

Regresión lineal: Cuando existe relación entre dos variables y podemos expresar esa relación en términos de una aproximación lineal (una recta que representa la relación entre las dos variables):

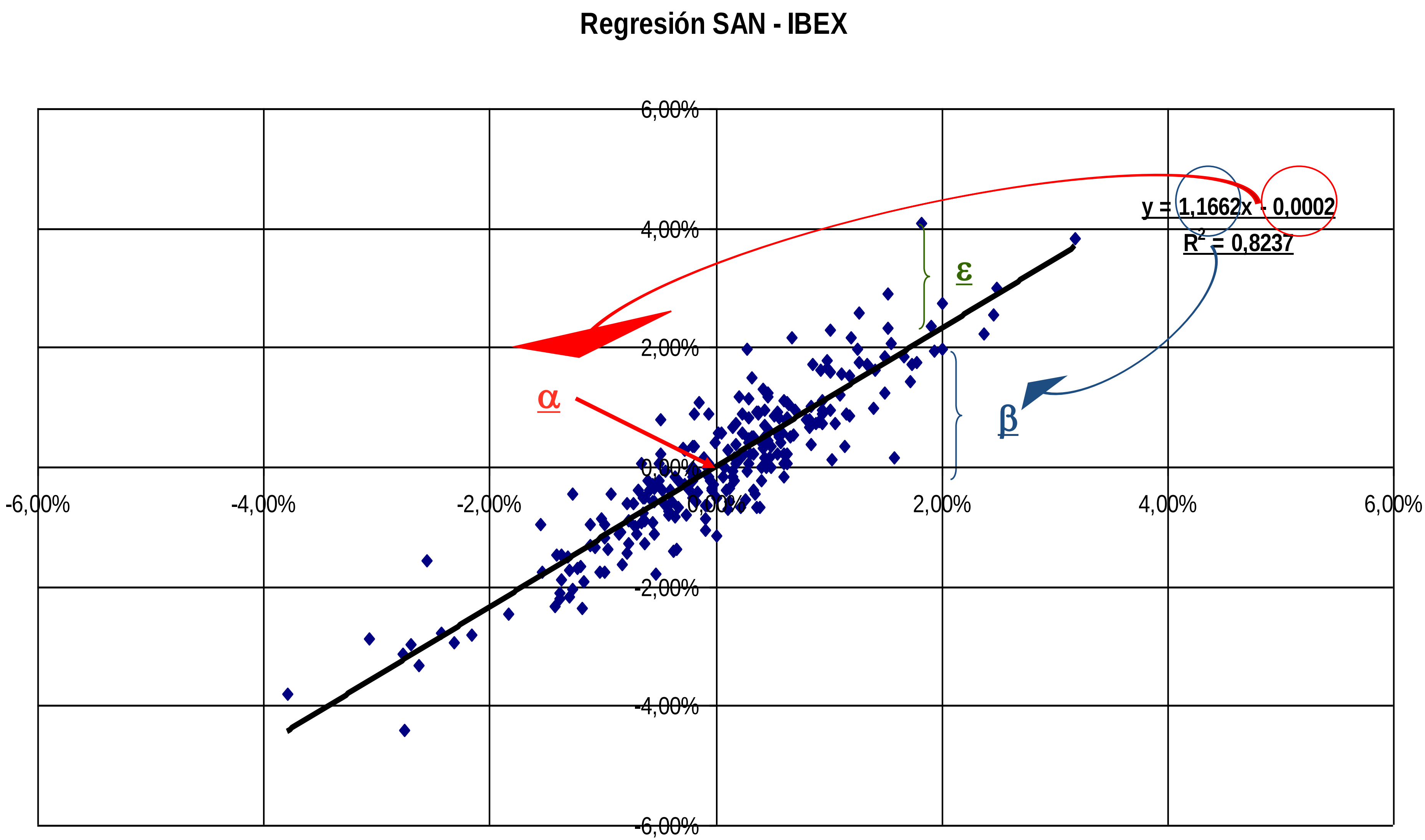
$$y_t = \alpha + \beta \cdot x + \varepsilon_t$$

El método que usaremos para la inferencia será el MCO (mínimos cuadrados ordinarios). Con este método, se representa la nube de puntos (valores reales) y, dibujando las infinitas rectas que pasan por el plano XY, hayamos la distancia de cada valor real a cada recta posible. Si sumamos todas esas distancias (elevadas al cuadrado para que no se compensen positivas con negativas) existirá una recta que minimiza todas esas distancias al cuadrado.

Esta aproximación lineal implica que para cada par de puntos existe un error e, que será exactamente la diferencia con respecto a la recta de regresión y que se intentará minimizar

$$\min = \sum_i (y_i - \alpha - \beta)^2 = \sum_i (\varepsilon_t)^2$$

Regresión SAN – IBEX 35



Regresión Lineal

Aplicando las condiciones de minimización llagamos a las siguientes expresiones:

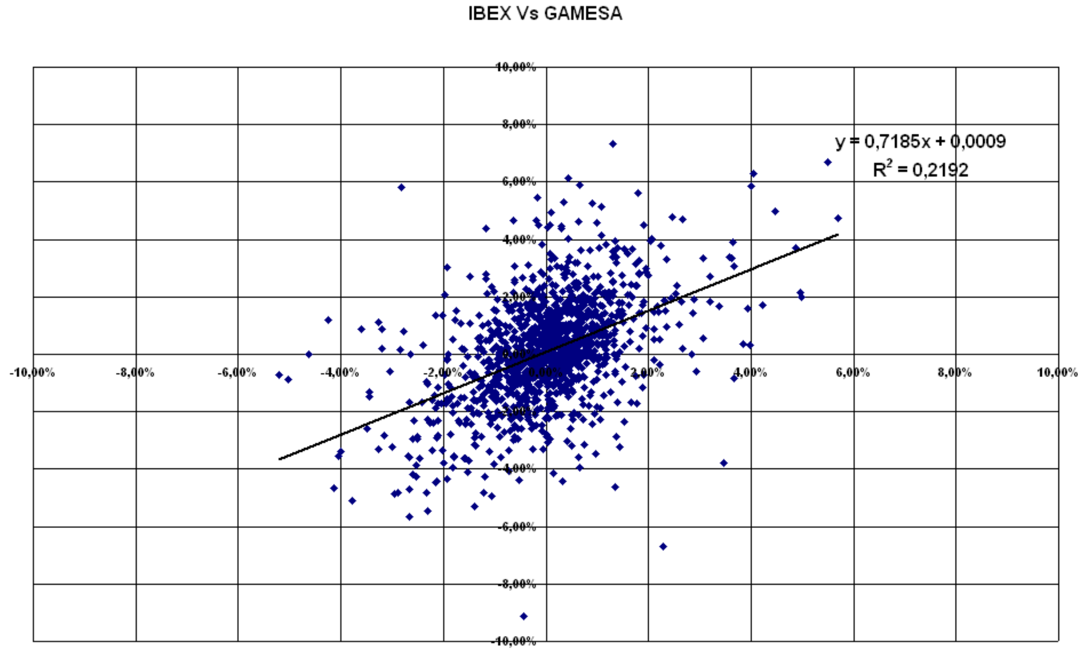
$$\beta = \frac{Cov(x,y)}{Var(x)} = \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} = \frac{S_{xy}}{S^2{}_x}$$

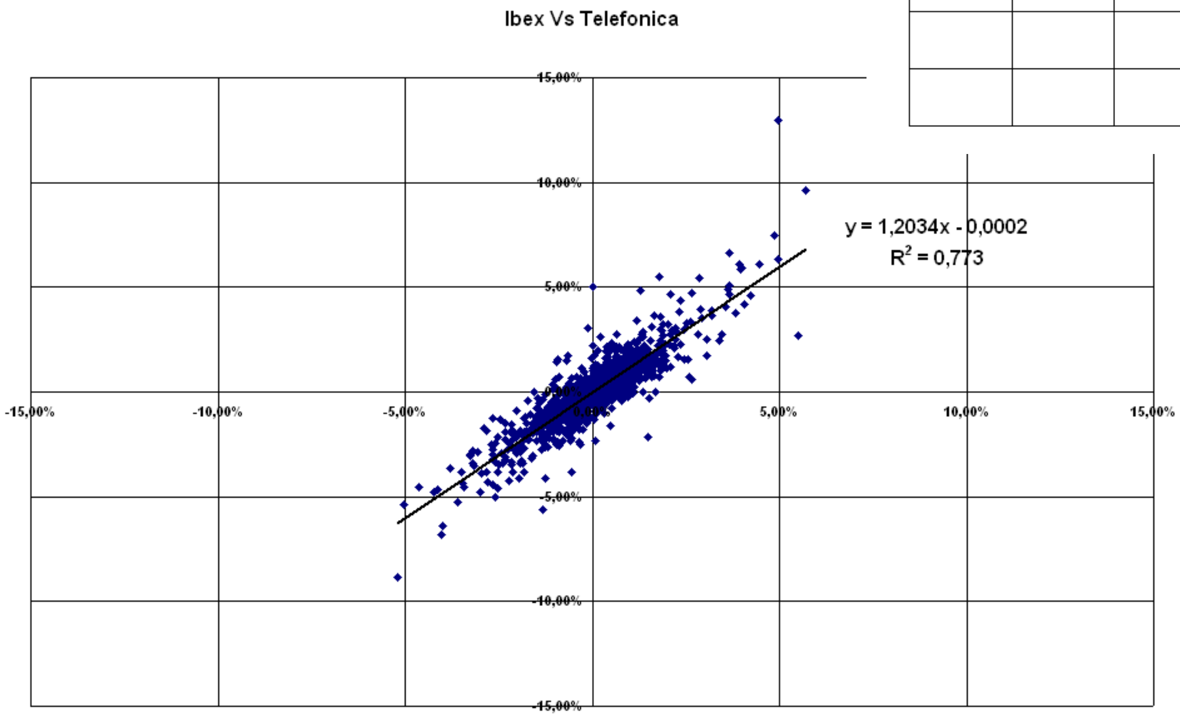
$$\alpha = \overline{y} - \beta \overline{x}$$

$$S^2{}_R = \frac{\sum \varepsilon_i^2}{n-2}$$ Estimador de la varianza residual.

Fiabilidad de las estimaciones

R2 es el coeficiente de bondad de ajuste e indica cómo de bien representa la recta de regresión al conjunto de datos. Es un valor entre 0 y 1, siendo 1 el ajuste perfecto (100%) y 0 el peor ajuste. Si la recta de regresión es representativa o no (fiabilidad de datos para estimaciones).





Fiabilidad de las regresiones. Bandas de confianza

Al disponer de distribuciones de los estimadores, podemos construir intervalos de confianza dentro de los cuales, y con un determinado nivel de probabilidad, se encontraran dichos parámetros.

Parámetro	α	β
Estimador	β_0	β_1
Desviación Típica estimadores	S_a	S_b
Bandas de confianza estimadas	β_0+/-t(n-2, α) x S_a	β_1+/-t(n-2, α) x S_b

Dónde:

t es la distribución T de Student con n-2 grados de libertad

α es el nivel de significación

$$S_a = \sqrt{\frac{\sum X_i^2}{n} \times \frac{1}{\sum\left(x_i - \overline{x}\right)^2} \times S_R^2}$$

$$S_b = \sqrt{\frac{S_R^2}{\sum\left(x_i - \overline{x}\right)^2}}$$

Fiabilidad de las estimaciones. Contraste de hipótesis

Con las bandas de confianza podemos observar cuáles son los rangos de variación de ambos estimadores, pero necesitamos algo que nos ayude tomar una decisión sobre los parámetros estimados.

β=b Para comprobar la hipótesis de que el coeficiente de regresión es igual a un valor determinado b usamos el siguiente estadístico:

$$t_b = \frac{\beta - b}{S_b}$$

Si $t_b < t$ entonces no se puede rechazar la hipótesis de que β=b, dónde t es una distribución de Student con n-2 grados de libertad para un nivel de significación especificado.

Excel. Función: estimación lineal

papeles | Fuente | Alineación | Número | Estilos

G5 f_x {=ESTIMACION.LINEAL(E3:E254;C3:C254;VERDADERO;VERDADERO)}

A	B	C	D	E	F	G	H	I
echa	IBEX		BBVA					
24/09/2007	14.494,70		15,78					
25/09/2007	14.311,90	-1,27%	15,53	-1,60%		1,13952618	-0,00014451	
26/09/2007	14.513,70	1,40%	15,60	0,45%		0,02868874	0,00048118	
27/09/2007	14.581,60	0,47%	15,65	0,32%		0,86321641	0,00762552	
28/09/2007	14.576,50	-0,03%	15,61	-0,26%		1577,70465	250	
01/10/2007	14.603,00	0,18%	15,72	0,70%		0,09174129	0,01453715	
02/10/2007	14.790,00	1,27%	16,24	3,25%				
03/10/2007	14.783,80	-0,04%	16,26	0,12%				

Se inserta la función en modo matricial en 5 filas x 2 columnas. Los datos que nos ofrece son los siguientes:

1,13952618 = β_1

0,02868874= Desvest de β_1

0,86321641= R2

1577,70465=Estadístico F

0,09174129=$\Sigma(y_t\text{-}y)^2$

-0,00014451= β_0

0,00048118= Desvest de β_0

0,00762552=Error estándar de la estimación=S_R

250=número de datos menos 2 grados de libertad

0,01453715=$\Sigma\varepsilon^2$

Documentación complementaria

Modelo de mercado de Sharpe

- Sharpe considera que el modelo de Markowitz y todos aquellos que utilizan la matriz de varianzas y covarianzas son difíciles de aplicar cuando se desean constituir carteras con un elevado número de activos, porque el número de parámetros a estimar es elevado.
- Por ello desarrolla un modelo de valoración y clasificación de activos, con el cual, a través de una regresión relaciona el rendimiento de un activo con el del mercado (o de un índice de referencia). El coeficiente de regresión es el coeficiente de Beta.

$$R_{it} = \alpha_i + \beta_i * R_{Mt} + \varepsilon_{it}$$

R_{it} indica la rentabilidad ofrecida por el título i en el período t.

R_{Mt} representa el rendimiento ofrecido por el mercado en el período t.

α_i como término independiente del modelo, expresa la parte del rendimiento del título i que es independiente del mercado.

β_i por su lado, indica cómo se comporta la rentabilidad del título i ante variaciones en el rendimiento del mercado.

ε_{it} es la perturbación aleatoria del modelo econométrico-financiero planteado por Sharpe. Su sentido se refiere a la parte de rentabilidad restante que no se explica por el modelo debida, por tanto, a otros factores no contemplados por el modelo.

CAPM

En el modelo CAPM desarrollado por Sharpe (posteriormente retocado por Lintner y Black) se establece que la rentabilidad de cualquier activo es función de la rentabilidad libre de riesgo, un riesgo sistemático dependiente del mercado y un riesgo diversificable correspondiente al propio activo.

$$Rc = rf + \beta(Rm - rf)$$

$$\text{Rc} = \text{Rentabilidad esperada de la cartera}$$

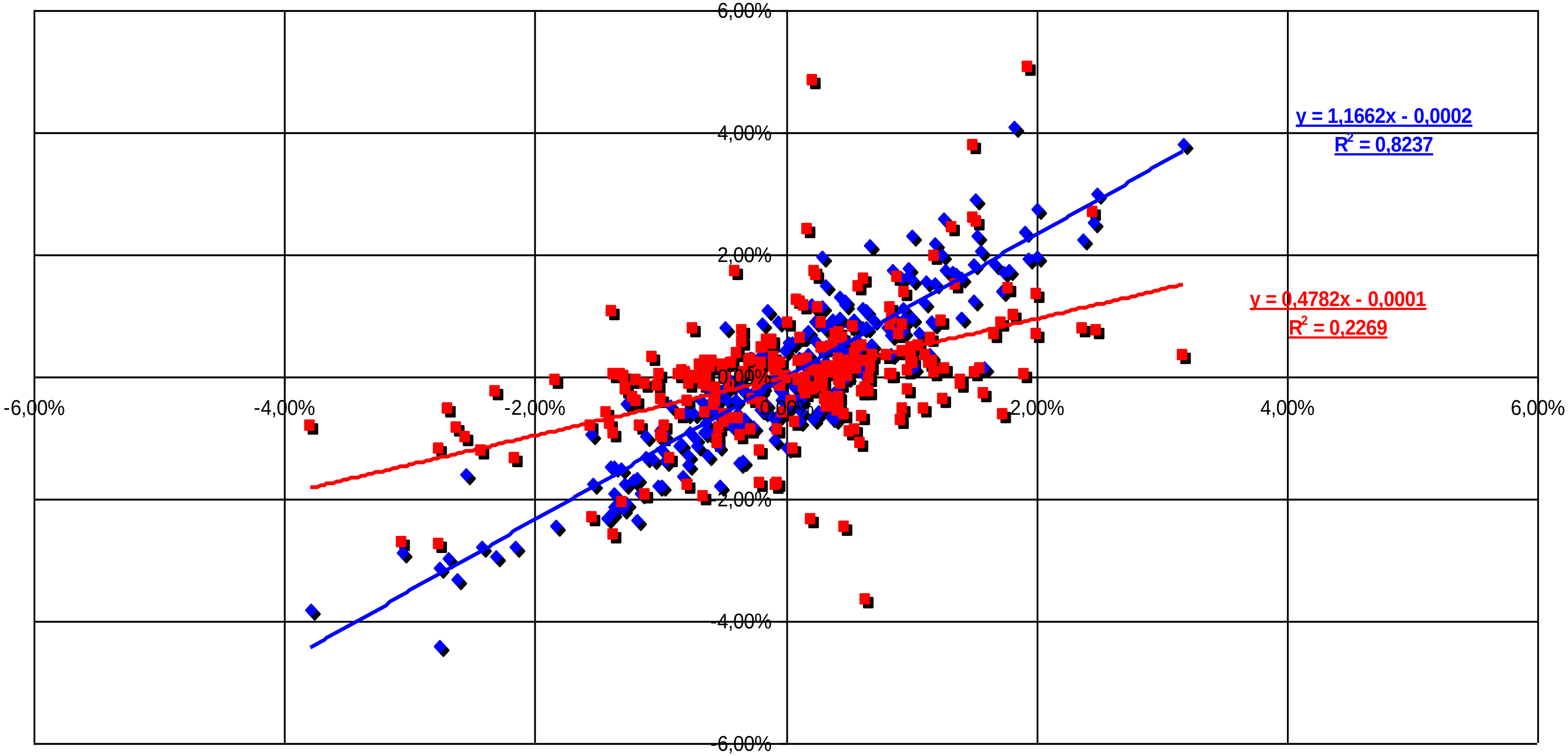
$$\text{rf} = \text{Rentabilidad Libre de riesgo}$$

$$\text{Rm} = \text{Rentabilidad del mercado}$$

$$\beta = \frac{\text{Covarianza (Rc, Rm)}}{\text{Varianza (Rm)}} \quad \text{(Beta de la cartera)}$$

Beta

Aunque este modelo presenta algunas limitaciones que hay que considerar…….



Beta. Coeficiente Beta, significado

$\beta = 1$ Los movimientos del índice y de la cartera son de igual intensidad.

$\beta > 1$ Es mayor el porcentaje de variación de la cartera que el del índice.

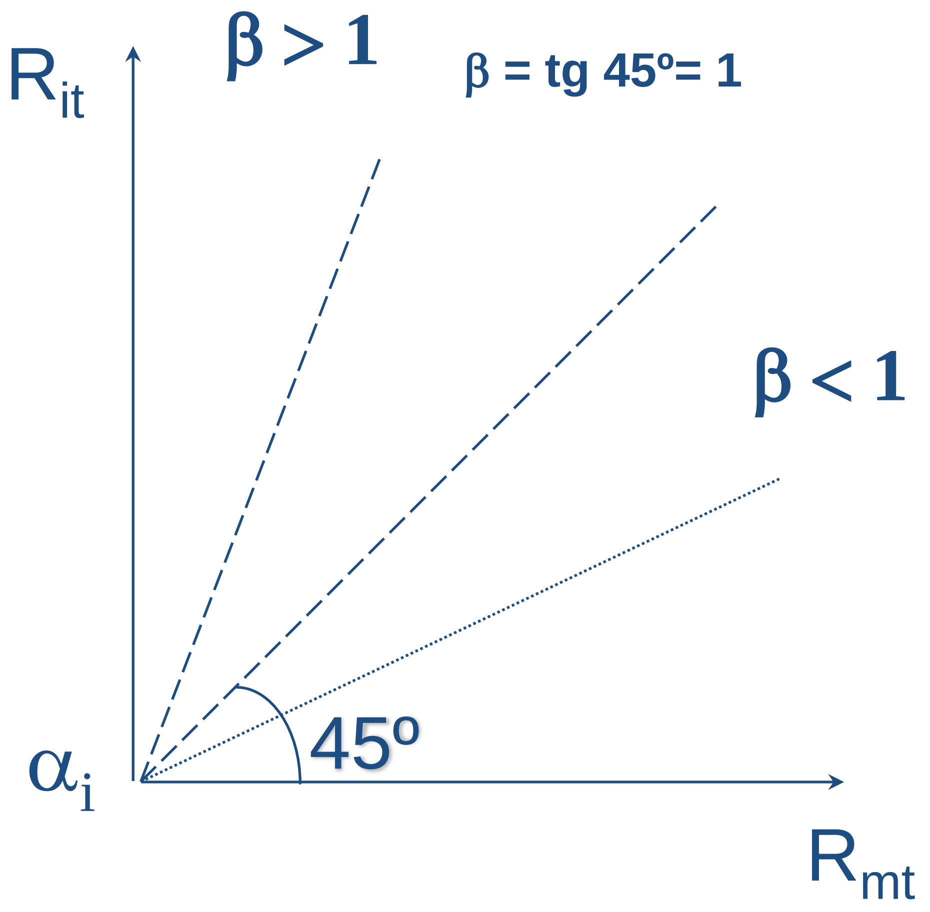
$0 < \beta < 1$ Es menor el porcentaje de variación de la cartera que el índice, por lo que en número de contratos necesarios para la cobertura será menor .

Beta. Coeficiente Beta, interpretación

$\beta < 1$ ➡ valores defensivos

$\beta = 1$ ➡ valores neutros

$\beta > 1$ ➡ valores agresivos



R_{it}= Rendimiento del Activo i en el momento t

R_{mt} = Rendimiento del mercado en el momento

Beta. Calculo de la Beta de una cartera

Sea una cartera compuesta por:

1.000 TEF,	Precio = 12,77	β = 1,20
1.500 BBVA,	Precio = 11,08	β = 1,47
2.000 ITX,	Precio = 16,00	β = 0,53
1.000 REE,	Precio = 13,24	β = 0,22

¿Cuál sería la Beta de la cartera?

$$\beta_c = \frac{X_1N_1\beta_1 + X_2N_2\beta_2 + X_3N_3\beta_3 + X_4N_4\beta_4}{X_1N_1 + X_2N_2 + X_3N_3 + X_4N_4} =$$

$$= \frac{1000\ x12.77\, x1.20 + 1500\ x11.08\, x1.47 + 2000\ x16.00\, x0.53 + 1000\ x13.24\, x0.22}{1000\ x12.77 + 1500\ x11.08 + 2000\ x16 + 1000\ x13.24} = 0.80$$

INSTITUTO BME

BME