

ESTADÍSTICA Y HERRAMIENTAS COMPUTACIONALES

Presentar elementos estadísticos, probabilísticos y herramientas necesarias para soportar decisiones financieras.

“There are three types of lies -- lies, damn lies, and statistics.”

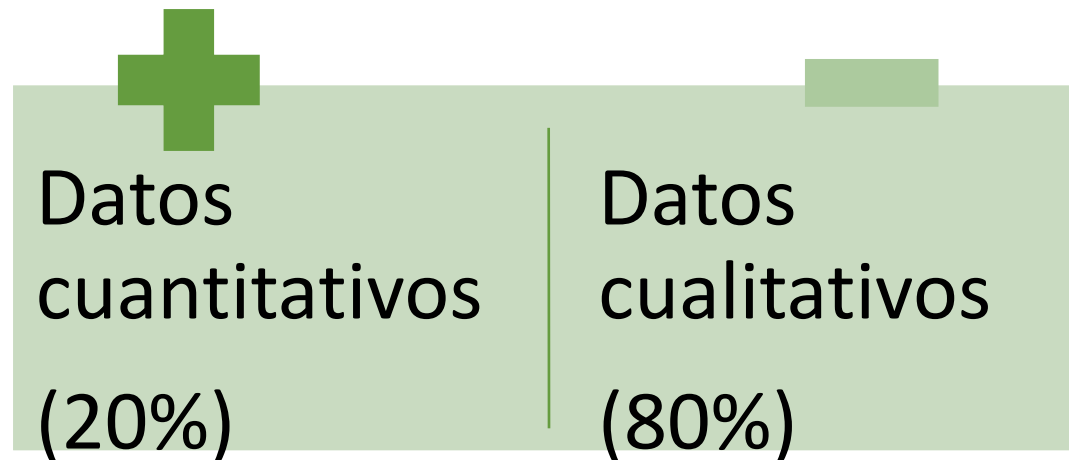
— **Benjamin Disraeli**

“Statistically speaking, there is a 65 percent chance that the love of your life is having an affair. Be very suspicious.”

— **Scott Dikkers**

Tipos de datos financieros

- Series de tiempo (x-axis: frecuencia – y-axis: precios, tasas, cantidades, etc.)
- Cross-sectional data (Datos transversales) – Caracterización de un estudio en un periodo de tiempo
- Panel data – es una serie de tiempo de un estudio en varios periodos del tiempo



Modelo de la naturaleza a las finanzas...

Distribución de probabilidad

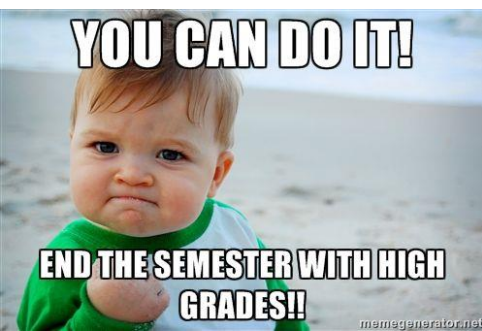
$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du$$

Densidad

$$\phi(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Louis Bachelier's "Theory of Speculation" Mark H. A. Davis, Imperial College pg. 2, 9, 12



Efficient Market Hypothesis (Teoría de los mercados eficientes)

“The weak form of the Efficient Market Hypothesis (EMH) asserts that prices fully reflect the information contained in the historical sequence of prices. Thus, investors cannot devise an investment strategy to yield abnormal profits on the basis of an analysis of past price patterns (a technique known as technical analysis). It is this form of efficiency that is associated with the term 'Random Walk Hypothesis'.

The semi-strong form of EMH asserts that current stock prices reflect not only historical price information but also all publicly available information relevant to a company's securities. If markets are efficient in this sense, then an analysis of balance sheets, income statements, announcements of dividend changes or stock splits or any other public information about a company (the technique of fundamental analysis) will not yield abnormal economic profits.

The strong form of EMH asserts that all information that is known to any market participant about a company is fully reflected in market prices. Hence, not even those with privileged information can make use of it to secure superior investment results. There is perfect revelation of all private information in market prices.”

Burton G. Malkiel, Efficient Market Hypothesis

Quienes son más complicados hombres o mujeres...

Women's Hormone Equations

$$\frac{dx_i}{dt} = x_i \{K - D(X - M_1 x_i)(X - M_2 x_i)\}, \quad i = 1, \dots, N, \quad \frac{1}{M_1} + \frac{1}{M_2} < 1, \quad X = \sum_{j=1}^N x_j.$$

Men's Hormone Equations

$$\begin{aligned} \lambda(t) &= H_{1,2} \left(\int_{(t-l_{1,2}) \vee 0}^{(t-l_{1,1}) \vee 0} X_{Te}(r) dr, \int_{(t-l_{2,2}) \vee 0}^{(t-l_{2,1}) \vee 0} X_G(r) dr \right), \quad \int_{s-l_2}^{s-l_1} X_A(r) dr \stackrel{\text{def}}{=} \begin{cases} \frac{1}{l_2 - l_1} \int_{s-l_2}^{s-l_1} X_A(r) dr, & \text{if } l_2 > l_1 \\ X_A(s - l_1), & \text{if } l_2 = l_1 \end{cases}, \\ p(s \mid T_G^{k-1}, \lambda(\cdot)) &= \gamma \times \lambda(s) \left(\int_{T_G^{k-1}}^s \lambda(r) dr \right)^{\gamma-1} \exp^{-\left(\int_{T_G^{k-1}}^s \lambda(r) dr \right)^\gamma}, \quad T_L^k = \left[\min_j \{T_G^j \mid T_G^j \geq T_L^{k-1} + \tau_L\} \right] + \tau_L, \\ N_G(t) &= \sum_{j=1}^{\infty} 1_{\{T_G^j \leq t\}}, \quad N_L(t) = \sum_{j=1}^{\infty} 1_{\{T_L^j \leq t\}}, \quad S_G(t) = H_3 \left(\int_{(t-l_{3,2}) \vee 0}^{(t-l_{3,1}) \vee 0} X_{Te}(s) ds \right) + \xi_G(t), \quad S_{Te}(t) = H_4 \left(\mu(t) \times \int_{(t-l_{4,2}) \vee 0}^{(t-l_{4,1}) \vee 0} X_L(s) ds \right) + \xi_{Te}(t), \\ S_L(t) &= H_{5,6} \left(\sum_{j=0}^{N_L(t)} \int_{(T_L^j - l_{5,2}) \vee 0}^{(T_L^j - l_{5,1}) \vee 0} X_G(s) ds \times \Gamma(t - T_L^j), \int_{(t-l_{6,2}) \vee 0}^{(t-l_{6,1}) \vee 0} X_{Te}(s) ds \right) + \xi_L(t), \\ d\xi_i(t) &= -\delta_i \xi_i(t) + \tau_i(S_i(t)) dB_i(t), \quad \xi_i(0) = 0, \quad \delta_i > 0, \quad i = Te, G, L, \\ A_G^j &= \int_{T_G^{j-1}}^{T_G^j} S_G(t) dt, \quad A_L^j = \int_{T_L^{j-1}}^{T_L^j} (1 - e^{-\eta(t - T_L^{j-1})}) S_L(t) dt, \quad M_i^j = \Psi_i(T_i^{j-1}, T_i^j) \times M_i^{j-1} + A_i^j, \quad i = G, L, \\ Z_G(t) dt &= [\beta_G + M_G^{N_G(t)} \psi_G(t - T_G^{N_G(t)})] dt, \quad dX_G(t) = \{-\alpha_G(X_G(t))X_G(t) + Z_G(t)\} dt + \sigma_G(X_G(t)) dW_G(t), \\ Z_L(t) dt &= [\beta_L + M_L^{N_L(t)} \psi_L(t - T_L^{N_L(t)}) + e^{-\eta(t - T_L^{N_L(t)})} + S_L(t)] dt, \quad dX_L(t) = \{-\alpha_L(X_L(t))X_L(t) + Z_L(t)\} dt + \sigma_L(X_L(t)) dW_L(t), \\ Z_{Te} dt &= [\beta_{Te} + S_{Te}(t)] dt, \quad dX_{Te}(t) = \{-\alpha_{Te}(X_{Te}(t))X_{Te}(t) + Z_{Te}(t)\} dt + \sigma_{Te}(X_{Te}(t)) dW_{Te}(t). \end{aligned}$$

Cresswell, C. (2014). Mathematics and sex.
TEDxSydney presentation.

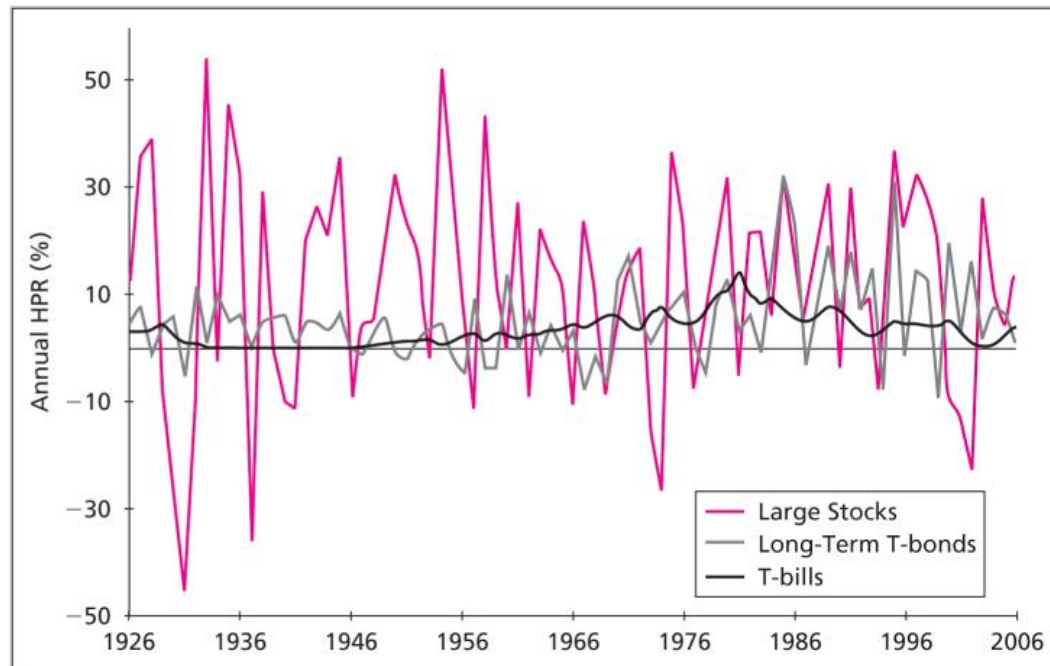
Estadísticas básicas

Media (μ)

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

Varianza ($\sigma^2 - s^2$) y desviación estándar ($\sigma - s$)

- $s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$
- $\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$



Valor esperado

Valor esperado

- $E(Y) = \sum_{i=1}^k P_i Y_i$

Varianza del valor esperado

- $Var(Y) = \sum_{i=1}^k P_i (Y_i - E(Y))^2$

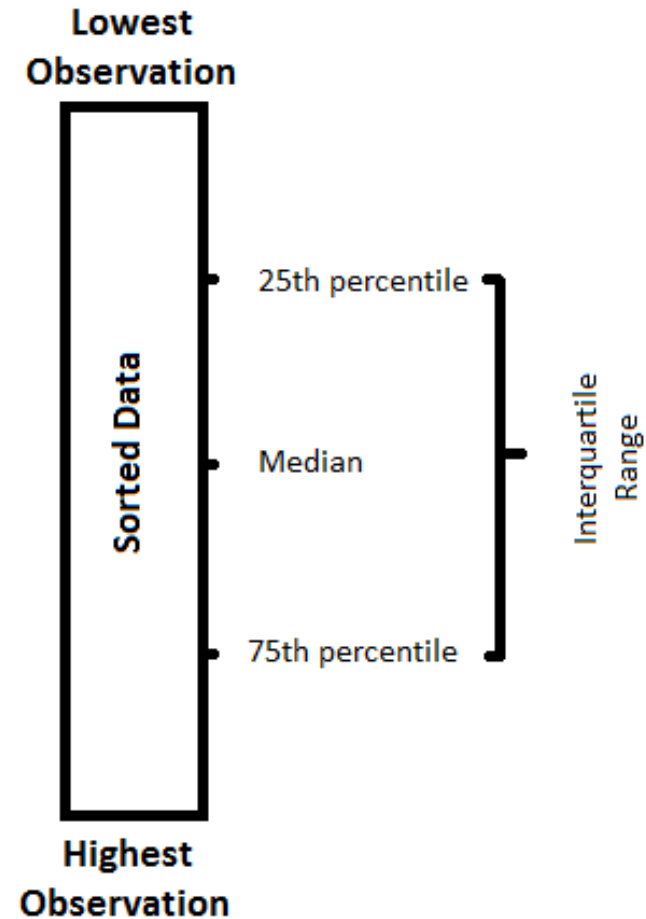
Escenario	Probabilidad	Indicador (Rentabilidad)
Boom	10%	5%
Normal	75%	2%
Recesión	15%	-2%

Estadísticas básicas

- Para una muestra de 5 observaciones en el mercado: {7, 2, 2, 4, 9}
- Para la **mediana** se ordenan los datos {2, 2, 4, 7, 9} y se toma el valor de la mitad, en caso de tener datos pares, se toman los centrales y se toma la media de los dos datos.
- **Rango** es la diferencia entre el máximo y mínimo $9-2=7$
- **Moda** es el dato que más se repite: 2

Rango intercuartil

- Se dividen los datos ordenados en 4/4. Los puntos de mayor interes son el percentil 25 y el percentil 75.
- El rango intercuartil es la diferencia entre estos dos y mide la dispersion de los datos.



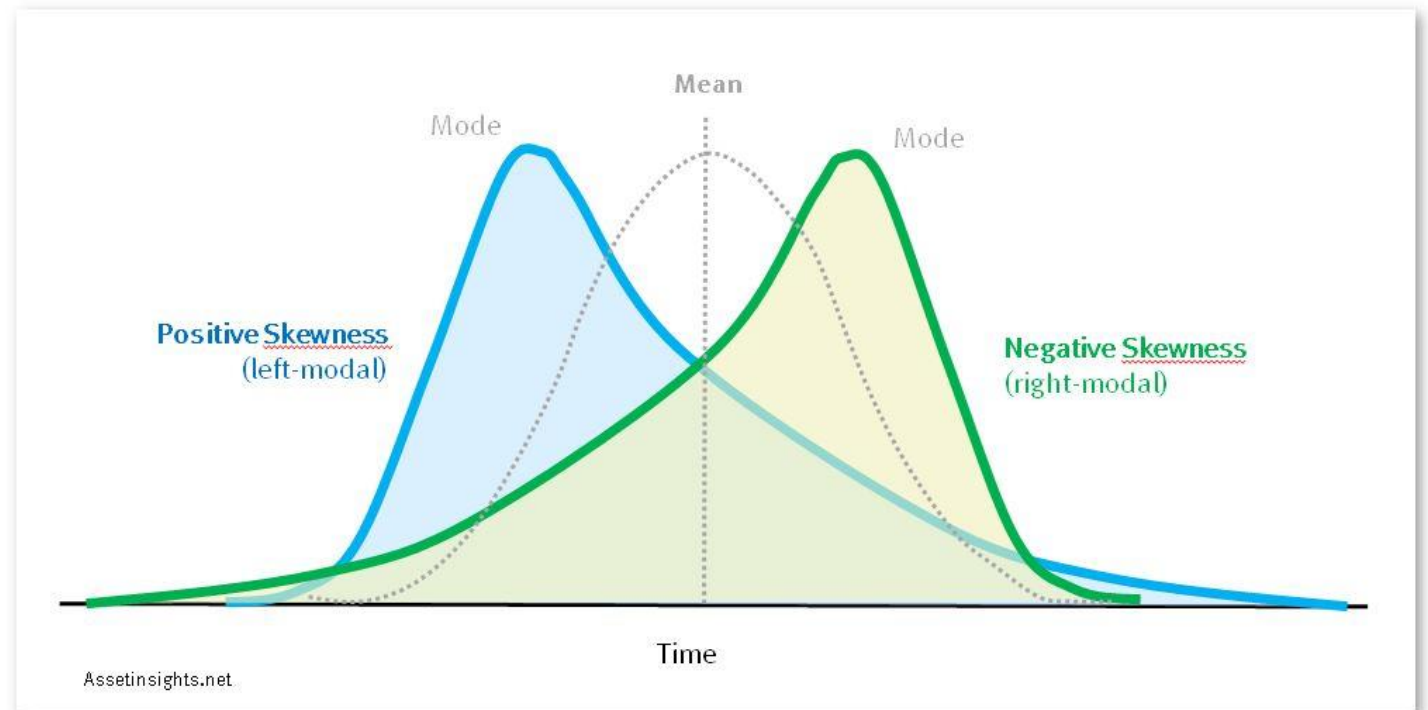
Gráficos



- Series de tiempo
- Diagramas de dispersión
- Distribución de frecuencia (Histogramas)
- **Distribución normal*

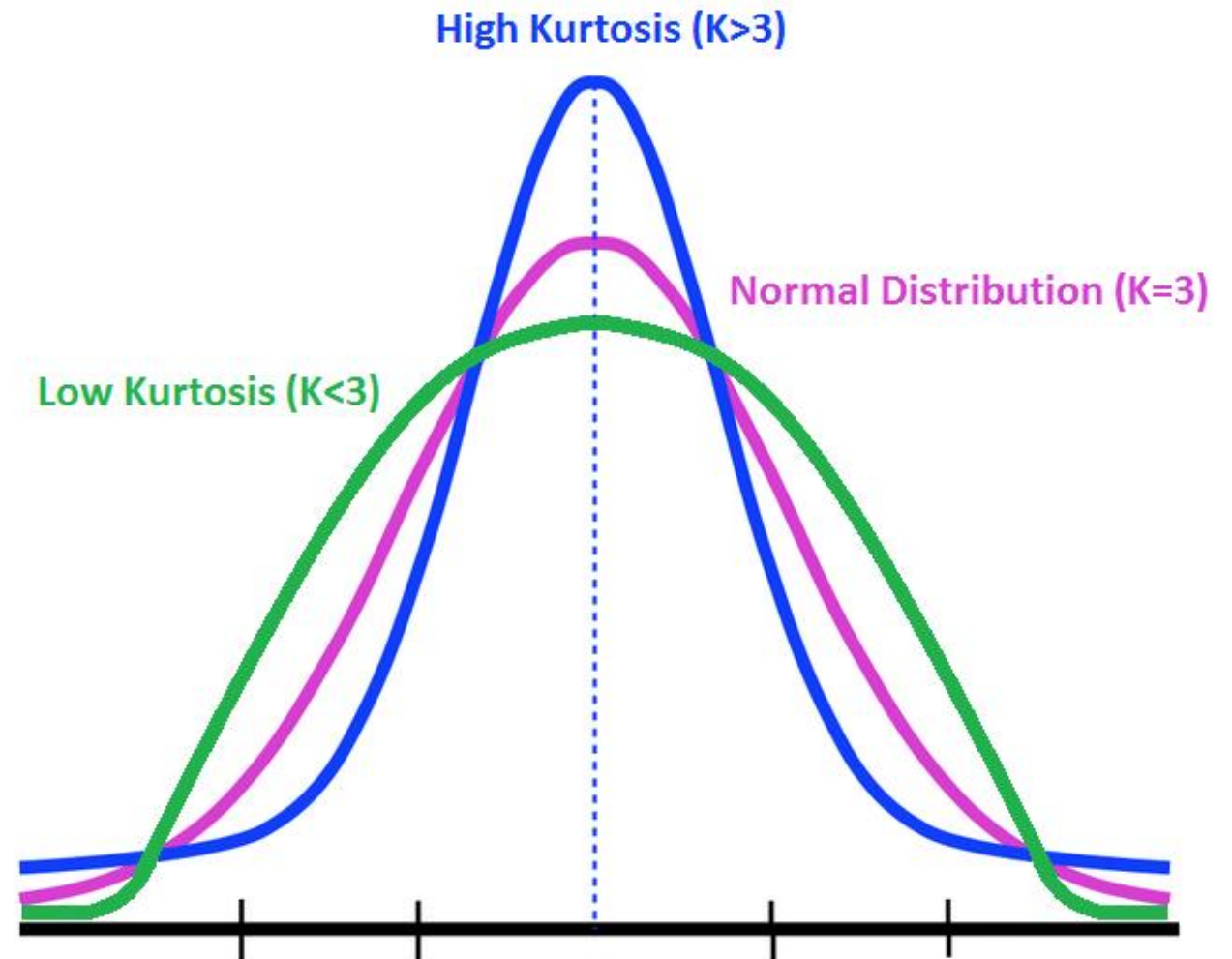
Skewness - Asimetria

$$Skewness = \frac{\frac{1}{n} \sum (x_i - \bar{x})^3}{\left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]^{3/2}}$$



Kurtosis - Curtosis

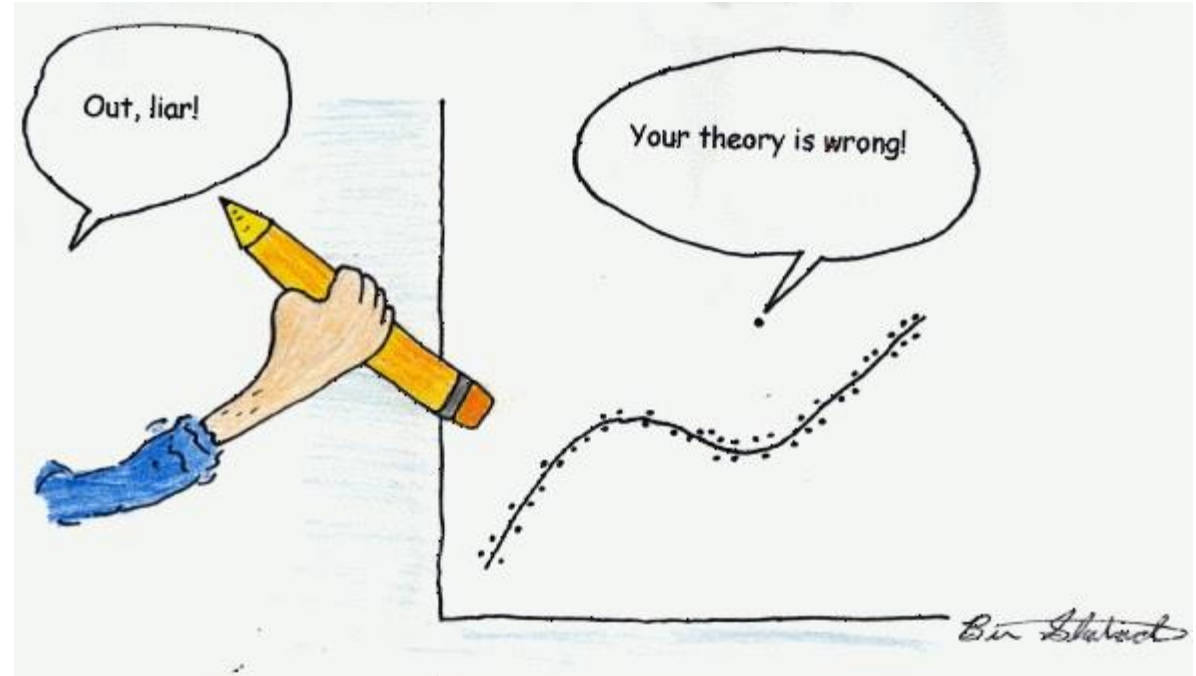
$$Kurtosis = \frac{\frac{1}{n} \sum (x_i - \bar{x})^4}{\left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]^2}$$



Outliers – Datos atípicos

Truncar: Excluir datos extremos, estamos borrando información importante.

Recortar (Winsorizing): Recortar a un percentil específico, todos los datos por encima y debajo de ese percentil toman el mismo valor.



Test para normalidad

Foco en Skewness y Kurtosis

- $JB = n \left[\frac{(Skewness)^2}{6} + \frac{(Kurtosis-3)^2}{24} \right]$
- $JB \sim \chi^2(2)$



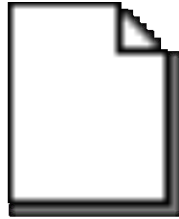
Carlos Jarque

Anil Bera

Test 1. Jarque-Bera – Validar si la distribución es normal

- Si el p-value es menor al 5% se rechaza la hipótesis que la distribución es normal.

Ejemplo Excel IQR, Skewness y Kurtosis (Bonus track Outliers)



Introducción Python

- [numpy](#): El popular paquete matemático de [Python](#), se utiliza tanto que mucha gente ya lo considera parte integral del lenguaje. Nos proporciona algunas funciones estadísticas que podemos aplicar fácilmente sobre los *arrays* de [Numpy](#).
- [scipy.stats](#): Este submodulo del paquete científico [Scipy](#) es el complemento perfecto para [Numpy](#), las funciones estadísticas que no encontremos en uno, las podemos encontrar en el otro.
- [statsmodels](#): Esta librería nos brinda un gran número de herramientas para explorar [datos](#), estimar modelos estadísticos, realizar pruebas estadísticas y muchas cosas más.
- [matplotlib](#): Es la librería más popular en [Python](#) para visualizaciones y gráficos. Ella nos va a permitir realizar los gráficos de las distintas distribuciones de datos.
- [seaborn](#): Esta librería es un complemento ideal de [matplotlib](#) para realizar gráficos estadísticos.
- [pandas](#): Esta es la librería más popular para análisis de [datos](#) y financieros. Posee algunas funciones muy útiles para realizar [estadística descriptiva](#) sobre nuestros datos y nos facilita sobremanera el trabajar con [series de tiempo](#).
- [pyMC](#): [pyMC](#) es un módulo de [Python](#) que implementa modelos estadísticos bayesianos, incluyendo la [cadena de Markov Monte Carlo\(MCMC\)](#). [pyMC](#) ofrece funcionalidades para hacer el análisis bayesiano lo mas simple posible.

Aplicaciones Python

- Usaremos el modulo de Jupyter Notebook

Data

- Aplicaciones de Bloomberg
- API Bloomberg
- Otras fuentes de información

Intro... Bloomberg

- Login
- Interfaz grafica y consultas basicas
- BMC Bloomberg Markets Concepts
- API Excel Bloomberg

Caso 1 – 25%

- Buscar una empresa local y una internacional con acción listadas, compararlas, resolviendo las siguientes preguntas...
- EJ. ESTE NO LO PUEDEN USAR COMPARAR BANCOLOMBIA VS BANK OF AMERICA

Guía

- **Preguntas:**

- De acuerdo con los gráficos que se pueden usar vistos en clase, que análisis pueden hacer de los dos activos, en cual invertirían y porque. Sustentar bien su respuesta.
- Que aspectos complementarios a los análisis estadísticos incluirían y dar un ejemplo con uno de ellos para esta comparación y análisis.
- Con base en los aspectos calculados (gráficos, tablas, datos calculados, etc.) si ustedes están asesorando a una persona con perfil de riesgo bajo que le diría respecto a estos dos activos.

Evaluación

- Examen plataforma virtual: 25%
- Caso 1: 25%
- Caso 2: 25%
- Seguimientos clases: 25%



(paréntesis)

- Retornos del periodo de tenencia

$$\text{Retorno } t = \frac{\text{Precio final} - \text{Precio inicial}}{\text{Precio inicial}}$$

- Retornos continuamente compuestos

$$\text{Retorno } t = \ln \frac{\text{Precio final}}{\text{Precio inicial}}$$

Distribuciones de frecuencia - Intro

Normalmente estos son los tipos de series de tiempo que utilizamos en la finanzas.

Que se tiene fecha y precio (u otras variables ejemplo cifras de estados financieros, indicadores, etc.)



Distribución de frecuencia - Histogramas

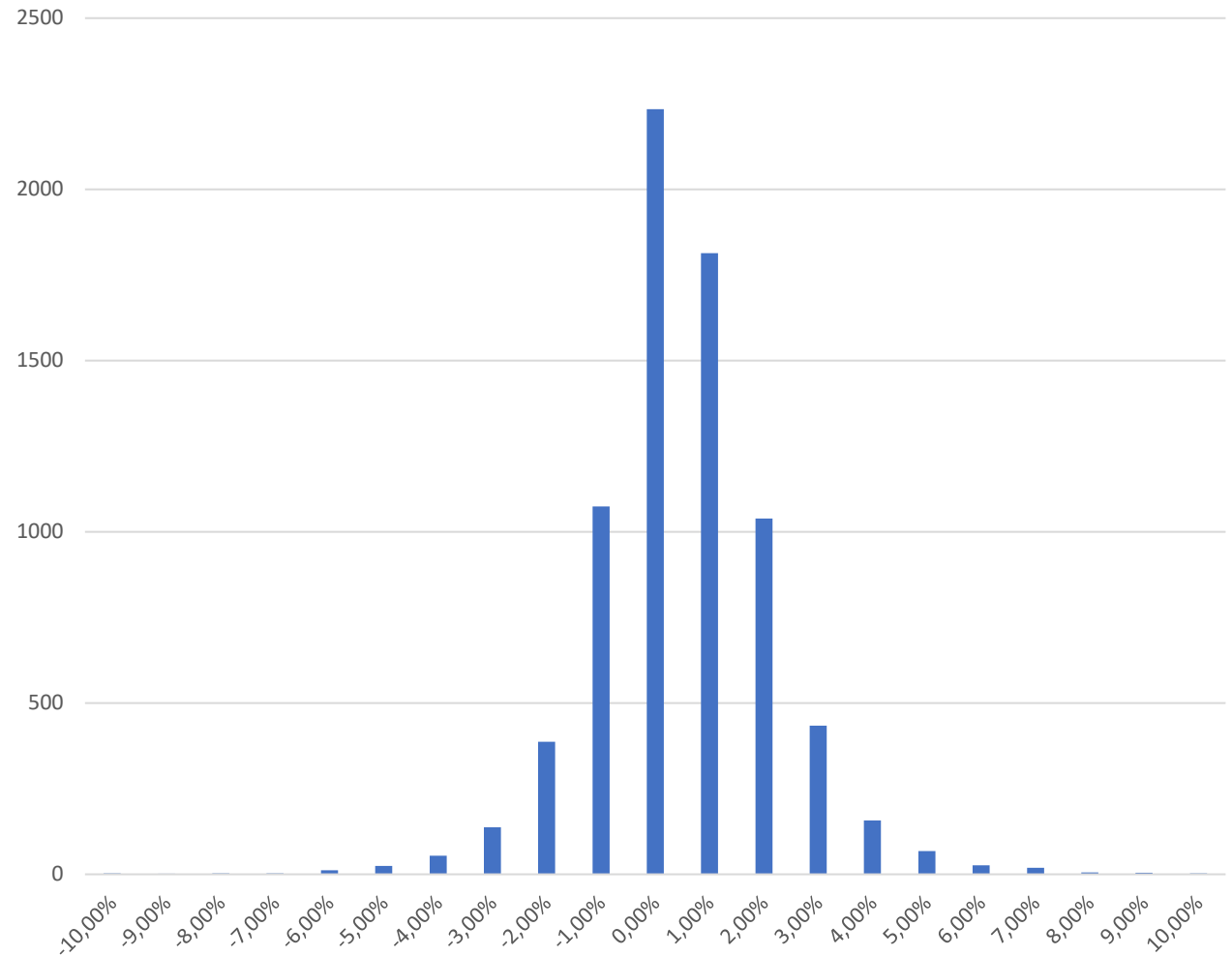


Frecuencia e histogramas

Resumen de datos con categorías definidas (ejemplo por rangos, intervalos, etc.)

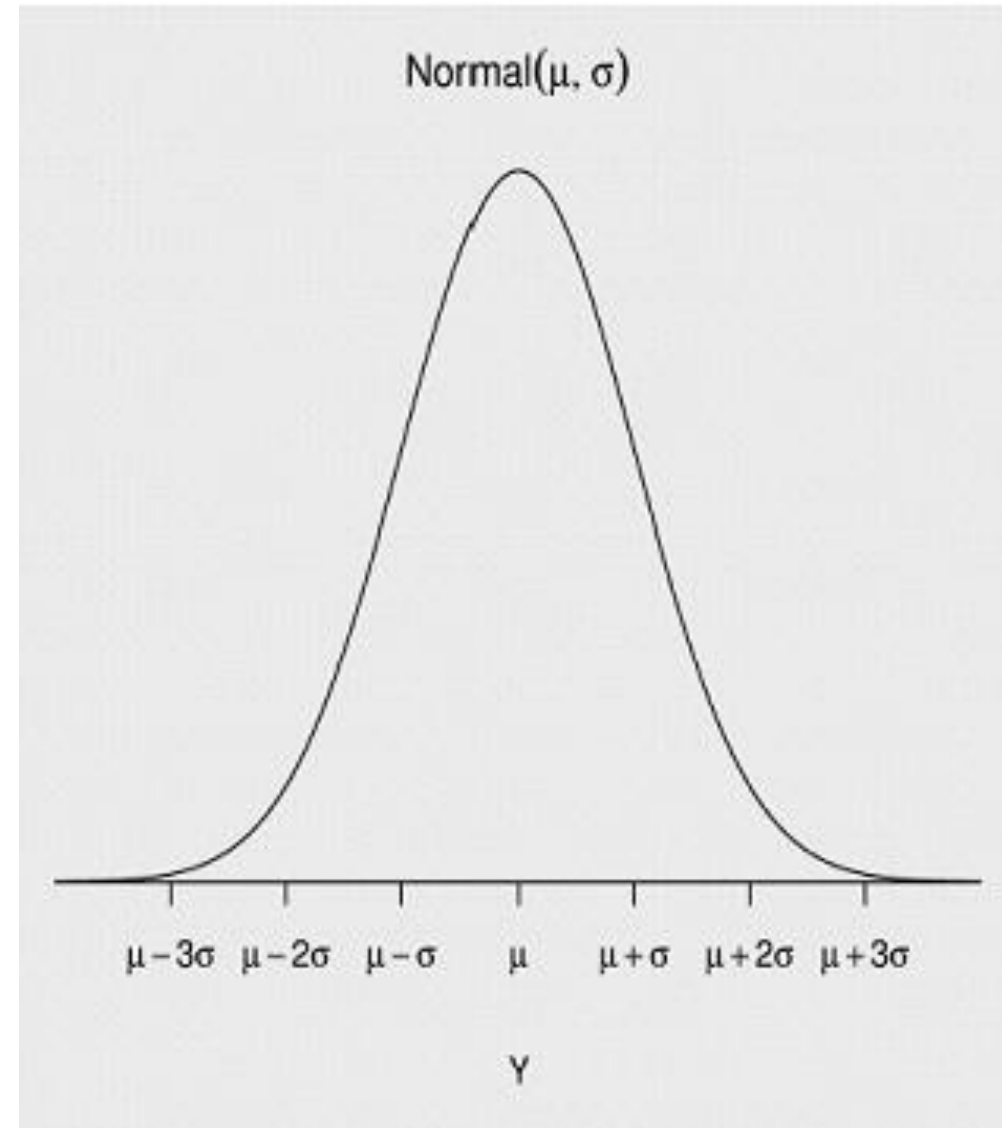
Se cuentan el número de observaciones en cada categoría.

El histograma muestra el comportamiento de la serie y cómo se analiza la información respecto a los datos que mas frecuencia tienen.

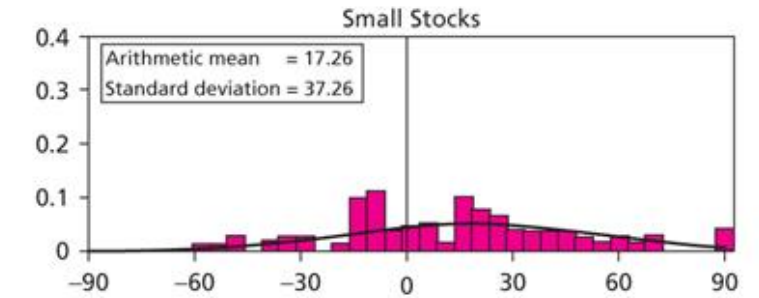
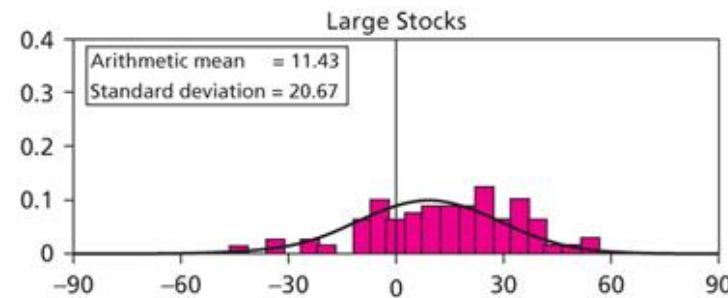
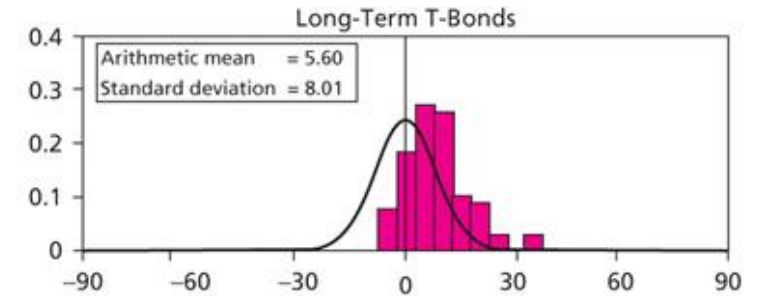
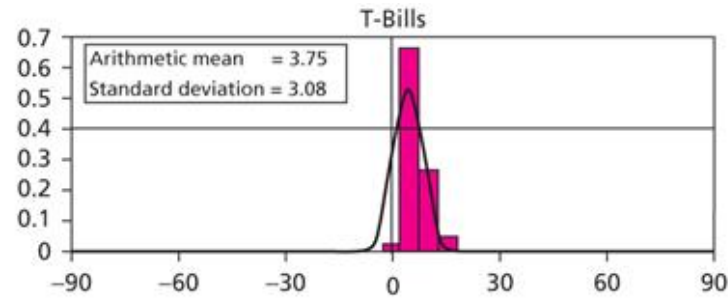


Distribución normal

- “Campana de Gauss”, una de las más importantes en Finanzas.
- Teóricamente es una de las gráficas que resume los datos financieros.
- Normalmente los modelos financieros asumen este tipo de distribuciones para usos varios (Machine Learning, AI, Predicciones, Regresiones)



Distribución normal



- Se caracteriza por estar definida la media μ y desviación estándar σ .
- Donde los datos $Y \sim N(\mu, \sigma)$ están distribuidos de esta forma.

[https://www.youtube.com/watch
?v=sxYrzzy3cq8](https://www.youtube.com/watch?v=sxYrzzy3cq8)

How statistics can be misleading - Mark Liddell



Caso de análisis Fondo Inversión

- Análisis de la empresa, validar resultados, interpretación económica y análisis de la decisión tomada de inversión.
- Se plantea caso hipotético de inversión en renta variable internacional, con perfil de riesgo alto y retorno esperado alto.
- Invertiría usted en esta empresa, sustentar su decisión.

¿Qué opinan?

- ¿Invertirían?
- ¿Cuánto?
- ¿A qué plazo?
- ¿Por qué?
- Sustentar con los datos generados

