ESTADÍSTICA APLICADA A LOS NEGOCIOS

12 de marzo de 2018

Dirección de Analítica Financiera
Sura Colombia
2018

INFORMACIÓN GENERAL

INFORMACIÓN GENERAL

1. Este es un curso teórico y aplicado.
2. No se requiere conocimiento previo en temas estadísticos.
3. Es necesario el uso de computador en el curso.
4. Idealmente deben llevar datos para usarlos en los espacios y aplicar los conceptos aprendidos.
5. Recomendación: Buscar conceptos básicos de manejo de RStudio.
6. ¡Ganas de aprender!

REFERENCIAS

REFERENCIAS

Referencias principales:

- Wackerly Others. (2010), *Estadística mantemática con aplicaciones* . Cengage.
- Casella G. & Berger R. (2002), *Statistical Inference*. Duxbury. Thomsom Learning.
- Walpole R. & Otros. (1998), *Probabilidad y estadística para ingenieros*. Pearson.
- Hull, J. (2009), *Risk management and financial institutions*. Wiley Sons.
- Evans & Olson. (2002). *Introduction to simulation and risk analysis*.
- Griffiths, W.; Hill, C. and Lim, G. (2011), *Principles of Econometrics*. (4th ed). New Jersey: Wiley.

FUENTES DE INFORMACIÓN

Nuestros datos son el laboratorio en el que buscamos aplicar todo lo que aprendamos.

HERRAMIENTAS TECNOLÓGICAS

SOFTWARE DE APOYO

- RStudio.
- Excel.

CONTENIDO

CONTENIDO

1. Estadística descriptiva.
2. Probabilidad.
3. Variables aleatorias discretas y continuas.
4. Inferencia estadística.
5. Distribuciones bivariadas.
6. Regresión lineal simple.
7. Regresión lineal múltiple.
8. Análisis multivariado.
9. Análisis de series de tiempo.

IMPORTANTE: DERECHOS

IMPORTANTE: DERECHOS

Este documento es el producto de la lectura, análisis y síntesis del contenido de los libros sugeridos como referencias principales, complementarias y otros.

INTRODUCCIÓN

INTRODUCCIÓN

Veamos algunas definiciones de estadística:

- "Some people hate the very name of statistics, but I find them full of beauty and interest". (Francis Galton).
- "The science of using information discovered from studying numbers". (Cambridge).
- "The practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample". (Oxford).

ESTADÍSTICA DESCRIPTIVA

MOTIVACIÓN

- ¿Qué haría si una de sus tareas a nivel laboral es obtener conocimiento a partir de un conjunto de datos?
- ¿Qué información a priori cree que necesitaría para hacer un adecuado análisis?
- ¿Qué conceptos estadísticos conoce o debería conocer para llevar a cabo un análisis de calidad?
- ¿Qué es para usted estadística descriptiva?

CONCEPTOS GENERALES

Siempre que se inicie un trabajo de investigación o aplicado se deben tener en cuenta los siguientes aspectos (Iral & Otros, 2009):

- Definición clara de objetivos.
- Obtención de datos.
- Análisis de datos.
- Informe de hallazgos.

CONCEPTOS GENERALES(2)

En la literatura existen múltiples definiciones del tema. Algunas son:

- A través de ella se pueden obtener medidas resúmenes de los datos por medio de funciones conocidas como estadísticos muestrales. (Iral & Otros, 2009).
- Es parte fundamental de cualquier análisis estadístico en la que se inicia la toma de decisiones que afectarán de manera significativa la investigación que se lleva a cabo. (Espejo & Otros, 2006).

CONCEPTOS GENERALES(3)

- **Población:**Es el conjunto de mediciones que se hace a un conjunto de individuos que tienen una característica común.
- **Muestra:**Es un subconjunto de la población.
- **Variable:**Característica que cambia de individuo a individuo o que puede cambiar en el tiempo. Ejemplos:
    1. Precio de cierre de la acción Preferencial de Bancolombia.
    2. Número de estudiantes inscritos a la Especialización en Finanzas.
    3. Ingreso salarial de un grupo de estudiantes de un curso de posgrado de una universidad específica.
    4. Gastos totales de una unidad de negocio en un año.
    5. Número de veces que el cliente ha entrado en default en un año.
    6. Número de días en mora que un cliente tiene en todas las obligaciones financieras.

CONCEPTOS GENERALES(4)

Las variables se pueden clasificar de la siguiente manera:

- Variable continua.
- Variable discreta.
- Variable categórica.

CONCEPTOS GENERALES(5)

Nivel de medición de las variables.

- Nominal: Se usa cuando la variable tiene codificación que lo asocia con una categoría.
- Ordinal: Se usa cuando se está haciendo referencia a algún tipo de jerarquía u orden.
- Intervalo: Se usa cuando la variable que se mide tiene como referencia un cero definido de manera arbitraria.
- Razón: Se usa cuando la variable que se mide tiene como referencia un cero absoluto.

CONCEPTOS GENERALES(6)

La estadística descriptiva puede tener dos orientaciones:

1. Análisis de datos no agrupados.
2. Análisis de datos agrupados.

ANÁLISIS DE DATOS AGRUPADOS

Algunos conceptos fundamentales para este tipo de análisis son:

- Comprender totalmente el fenómeno que se está investigando.
- Definir Número de clases o intervalos en los que se deben agrupar los datos.
  $k = 1 + 3{,}33 \log_{10}(n)$ o $k = \sqrt{n}$
- Hallar los valores máximo,x_{max} y mínimo, x_{min}.
  Recomendación:
  - Si los datos son enteros sume $0{,}5$ a x_{max} y reste $0{,}5$ a x_{min}.
  - Si los datos tiene un dígito decimal sume $0{,}05$ a x_{max} y reste $0{,}05$ a x_{min}.
- Calcular el rango ampliado, $R^* = x_{max} - x_{min}$.
- Calcular la amplitud. $A = \frac{R^*}{k}$

ANÁLISIS DE DATOS AGRUPADOS

- Calcular para cada intervalo o clase el valor medio, x_i.
- Calcular las siguientes frecuencias.
  * Frecuencia absoluta, f_i.
  * Frecuencia acumulada, F_i.
  * Frecuencia relativa, h_i.
  * Frecuencia relativa acumulada, H_i.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS

Existen diferentes tipos de medidas que permiten obtener información de los datos:

1. Medidas de posición y tendencia central.
2. Medidas de dispersión.
3. Medidas de forma.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (2)

Medidas de posición y tendencia central

- Media muestral: Busca identificar el centro de los datos.

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

- Moda: Es el dato que más se repite. En el caso de datos agrupados es la marca de clase, x_i, asociada con la mayor frecuencia absoluta.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (3)

Existen otras medias.

- Media geométrica.
- Media recortada.
- Media de Windsor.
- Trimedia.

TAREA: Consultar cada una y construir un ejemplo en el que tenga validez aplicarla.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (4)

Veamos a continuación las medidas de posición:

- Percentil: Medida de posición que permite identificar la proporción de observaciones que están por encima o por debajo de dicho percentil.

$$P_b = L + \frac{\left(\frac{nb}{100} - a\right) * A}{f}$$

  * b:Percentil que se desea hallar.
  * L:Límite inferior del intervalo que contiene al percentil.
  * n:Tamaño de la muestra.
  * a:Frecuencia absoluta acumulada del intervalo inmediatamente anterior al que contiene el percentil.
    NOTA: En caso que el percentil esté en el primer intervalo entonces $a = 0$.
  * A:Amplitud del intervalo.
  * f:Frecuencia absoluta del intervalo que contiene al percentil.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (5)

- Mediana: Es el valor que divide a la muestra en dos proporciones iguales, $50\,\%$ superior y $50\,\%$ inferior. Es denotada como, $\tilde{x}$. Concretamente $\tilde{x} = P_{50}$

¿Qué es un cuartil, un decil, un quintil?

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (6)

Medidas de dispersión

- Varianza: Mide qué tan alejados están los datos de la media.

$$S^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n}$$

- Desviación estándar: Es la raíz cuadrada positiva de la varianza.

$$\sigma = +\sqrt{S^2}$$

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (7)

- Rango intercuartil: Mide qué tan dispersos están el $50\,\%$ de los datos.

$$IQR = P_{75} - P_{25}$$

- Coeficiente de variación: Es el cociente entre la desviación típica y el valor absoluto de la media.

$$CV = \frac{\sigma}{|\bar{x}|}$$

  * Medida adimensional.
  * Permite comparar la dispersión de diferentes grupos o distribuciones.
  * Se usa frecuentemente en análisis de tipo financiero para comparar proyectos.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (8)

Medidas de Forma

- Sesgo o coeficiente de asimetría: Denotado por A, permite identificar el grado de asimetría de la distribución respecto a la media. Puede ser:
    * Simétrica: La distribución de los datos es simétrica respecto a la media. $A = 0$.
    * Sesgada a izquierda: Si $A < 0$, la distribución de los datos es sesgada a izquierda. Encontramos valores extremos en la izquierda del gráfico de distribución.
    * Sesgada a derecha: Si $A > 0$, la distribución de los datos es sesgada a derecha. Encontramos valores extremos en la derecha del gráfico de distribución.

MEDIDAS RESUMEN PARA DATOS AGRUPADOS (9)

- Curtosis: Denotado por K. Permite identifica el grado de apuntalamiento de la curva, es decir, qué tan concentrados están los datos centrados respecto a la media. Puede ser:
  * Platicúrtica. $K < 3$, implica que existe gran dispersión de los datos respecto a la media, es decir, la distribución de los datos es en forma aplanada.
  * Mesocúrtica. $K = 3$, este comportamiento es el de la distribución normal.
  * Leptocúrtica. $K > 3$, implica que existe gran concentración de los datos respecto a la media, es decir, la distribución de los datos es de colas más pesadas que las de la distribución normal.

  TAREA: Consultar qué es el exceso de curtosis, cómo se define y qué aplicaciones tiene en finanzas.

GRÁFICOS PARA EXPLORACIÓN DE DATOS

Existen diferentes tipos de gráficos que permiten identificar aspectos de interés en un conjunto de datos, tales como:

- Dispersión.
- Simetría.
- Tendencias.
- Forma funcional.

GRÁFICOS PARA EXPLORACIÓN DE DATOS (2)

Algunos gráficos dependiendo el tipo de análisis:

Análisis univariado

- Histograma.
- Polígono.
- Ojiva.
- Diagrama de tallos y hojas.
- Diagrama de puntos.
- Box Plot.

GRÁFICOS PARA EXPLORACIÓN DE DATOS (3)

Análisis bivariado

- Diagrama de dispersión.
- Diagrama de dispersión y gráfico de puntos.

Análisis multivariado

- Matriz de dispersión.
- Gráfico de SPIN.
- Gráfico de estrellas.
- Gráfico de rayos solares.
- Gráfico de caras.

MEDIDAS DE RELACIÓN ENTRE VARIABLES

Covarianza: Permite definir si existe asociación lineal entre dos variables.

$$\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- Si $\operatorname{cov}(x, y) > 0$ se dice que existe relación lineal positiva entre x e y.
- Si $\operatorname{cov}(x, y) < 0$ se dice que existe relación lineal negativa entre x e y.
- Si $\operatorname{cov}(x, y) = 0$ se dice que no existe relación lineal entre x e y.

MEDIDAS DE RELACIÓN ENTRE VARIABLES (2)

Correlación: Permite definir la fortaleza de la asociación lineal entre dos variables.

$$corr\,(x,y) = \rho_{xy} = \frac{\text{cov}(x,y)}{S_x S_y}$$

- Si $\rho_{xy} = 1$, relación lineal perfecta entre x e y.
- Si $\rho_{xy} = -1$, relación lineal perfecta-inversa entre x e y.
- Si $0{,}9 \leq \rho_{xy} \leq 1$, relación lineal positiva muy fuerte entre x e y.
- Si $0{,}8 \leq \rho_{xy} \leq 0{,}9$, relación lineal positiva fuerte entre x e y.
- Si $0{,}6 \leq \rho_{xy} \leq 0{,}8$, relación lineal positiva moderada x e y.
- Si $0{,}3 \leq \rho_{xy} \leq 0{,}6$, relación lineal positiva débil x e y.
- Si $-0{,}3 \leq \rho_{xy} \leq 0{,}3$, posiblemente no existe relación lineal x e y.