

Modelos Mixtos

“Definiciones básicas”

M.Sc. Henry Luis López García
Universidad Nacional Autónoma de Nicaragua, Managua
Facultad de Ciencia e Ingeniería



@Hen1985

“Los datos son el nuevo petróleo. Es valioso, pero si no está refinado, realmente no se puede usar. Tiene que cambiarse a gas, plástico, productos químicos, etc. para crear una entidad valiosa que impulse la actividad rentable; entonces los datos deben desglosarse, analizarse para que tengan valor ”.

Sin datos empíricos respecto de la situación actual, no es posible trazar con confianza nuestro camino hacia el cumplimiento de los Objetivos de Desarrollo Sostenible. Con ese propósito, el presente informe también refleja los desafíos enfrentados en la recopilación, el procesamiento, el análisis y la diseminación de la información suficientemente desglosada, accesible, oportuna y confiable y exige una mejor formulación de políticas basada en los datos. Las tecnologías de hoy en día posibilitan la recolección de los datos necesarios para cumplir con la promesa de no dejar a nadie atrás, pero es imprescindible el liderazgo político, los recursos y el compromiso de utilizar las herramientas disponibles en la actualidad (ONU, 2018).

Estadística

- Objetivos bien definidos
- Presupuesto (\$)
- Recursos (humanos y materiales)
- Prueba Piloto
- Variable de diseño
- Tamaño de la muestra

Planificación



- Aplicación de los formularios a través de la técnica por encuesta
- Extracción de datos (redes sociales, IoT)
- Registros vitales
- CENSO

Recolección



- Almacenamiento
- Limpieza de datos
- Análisis Exploratorio (EDA)
- Aplicación estadística paramétrica o no paramétrica
- Modelización

Procesamiento

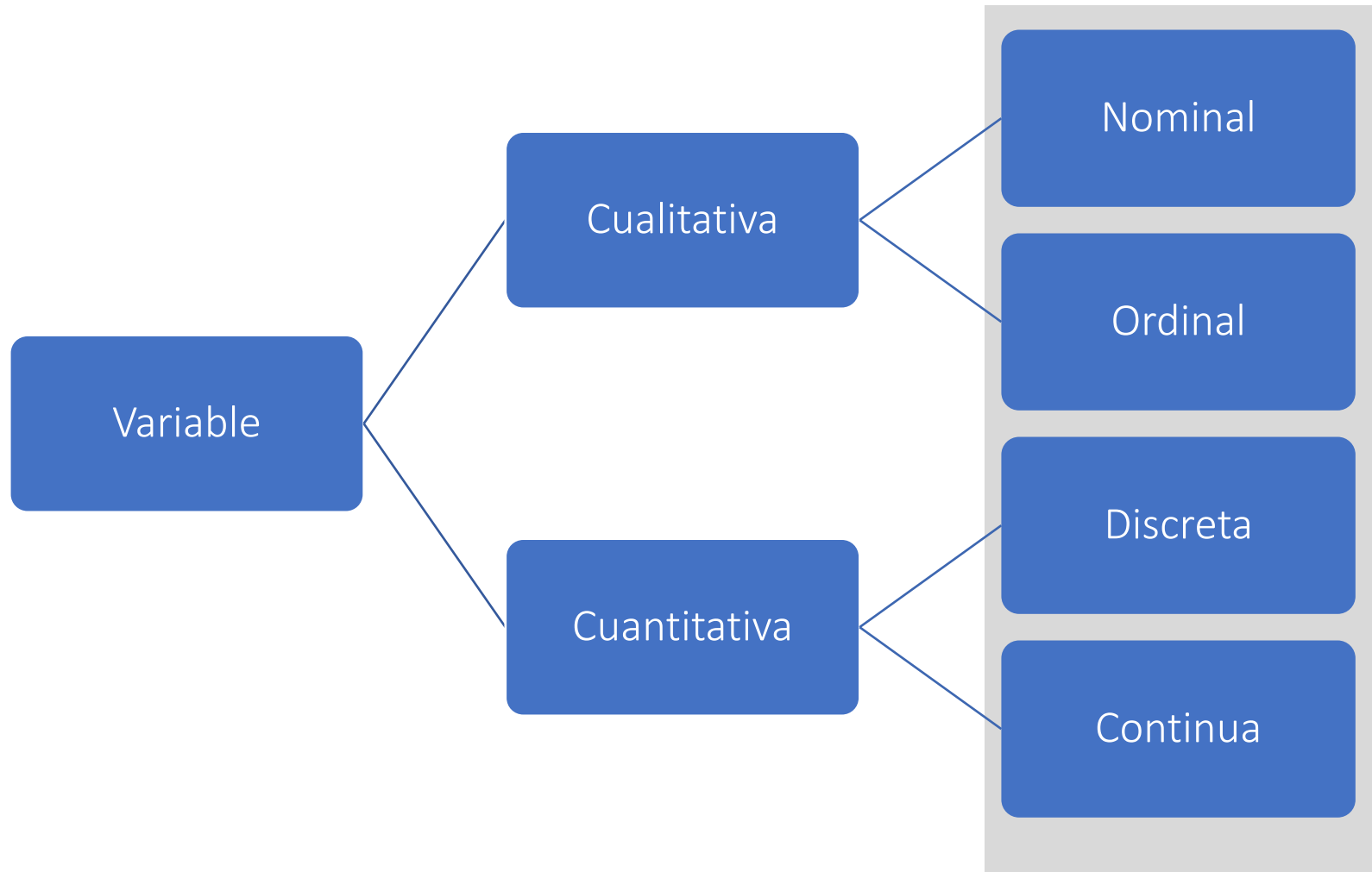


- Análisis de los resultados
- Gobernanza
- Line base
- Medición de impacto
- Mejora continua

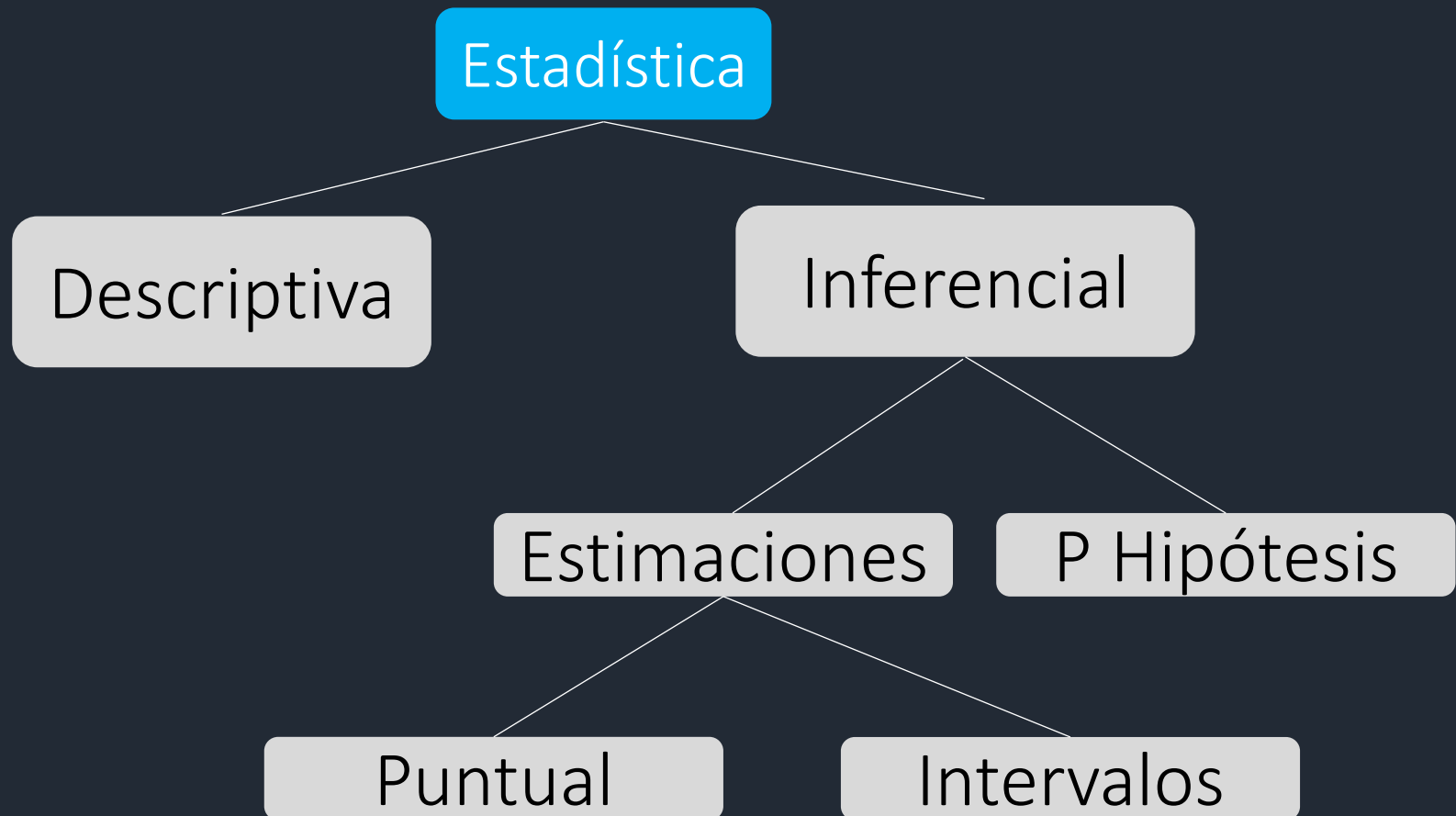
Análisis



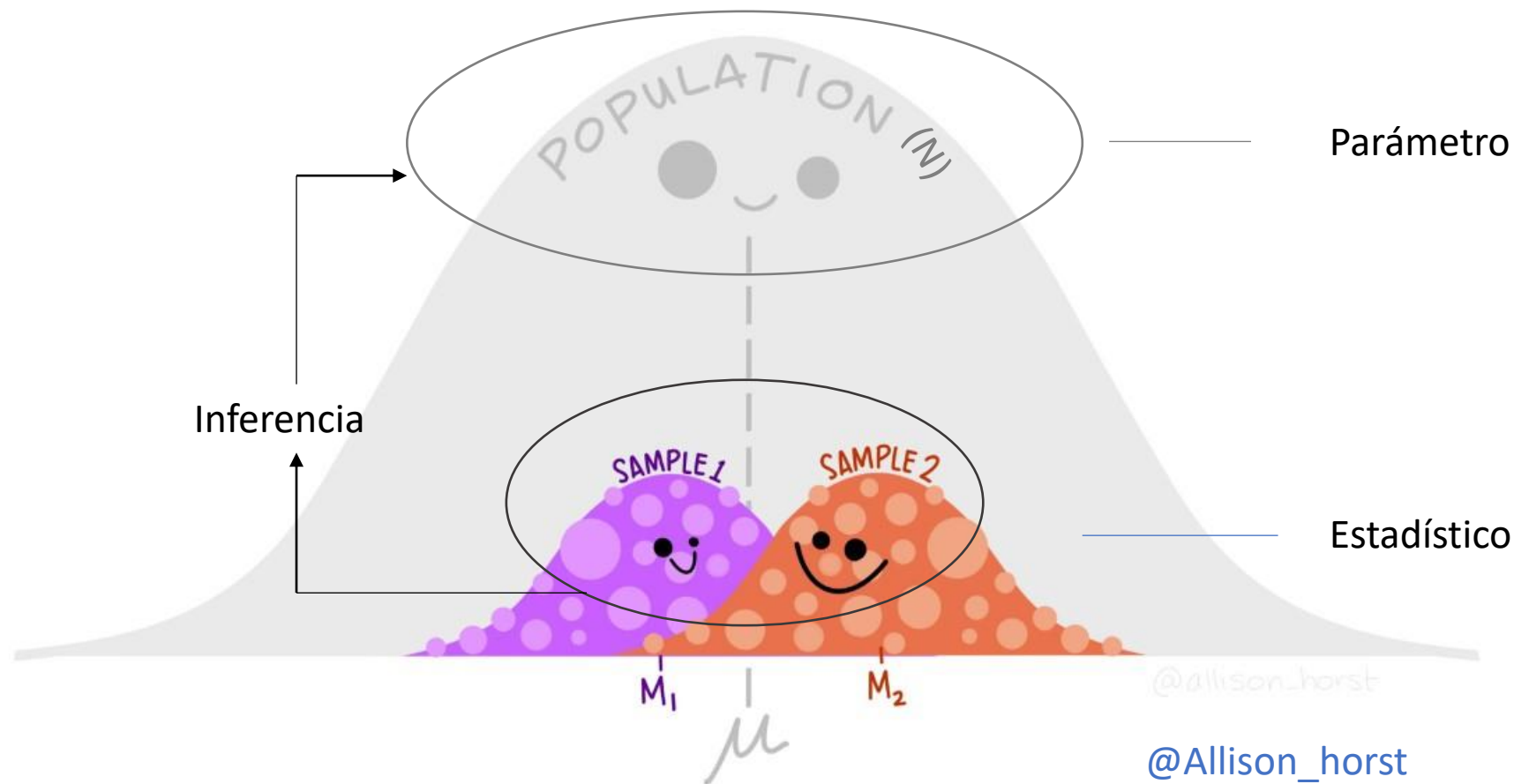
Variables



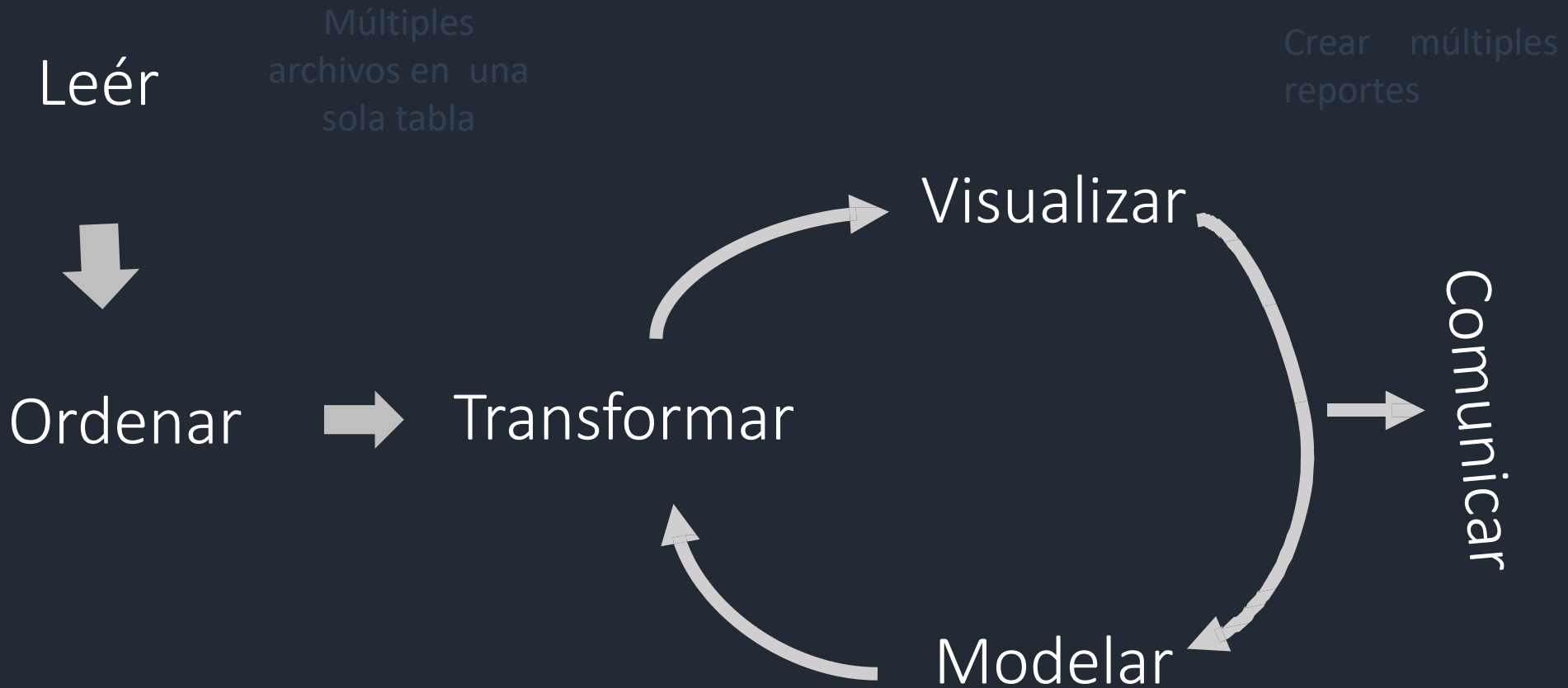
División clásica



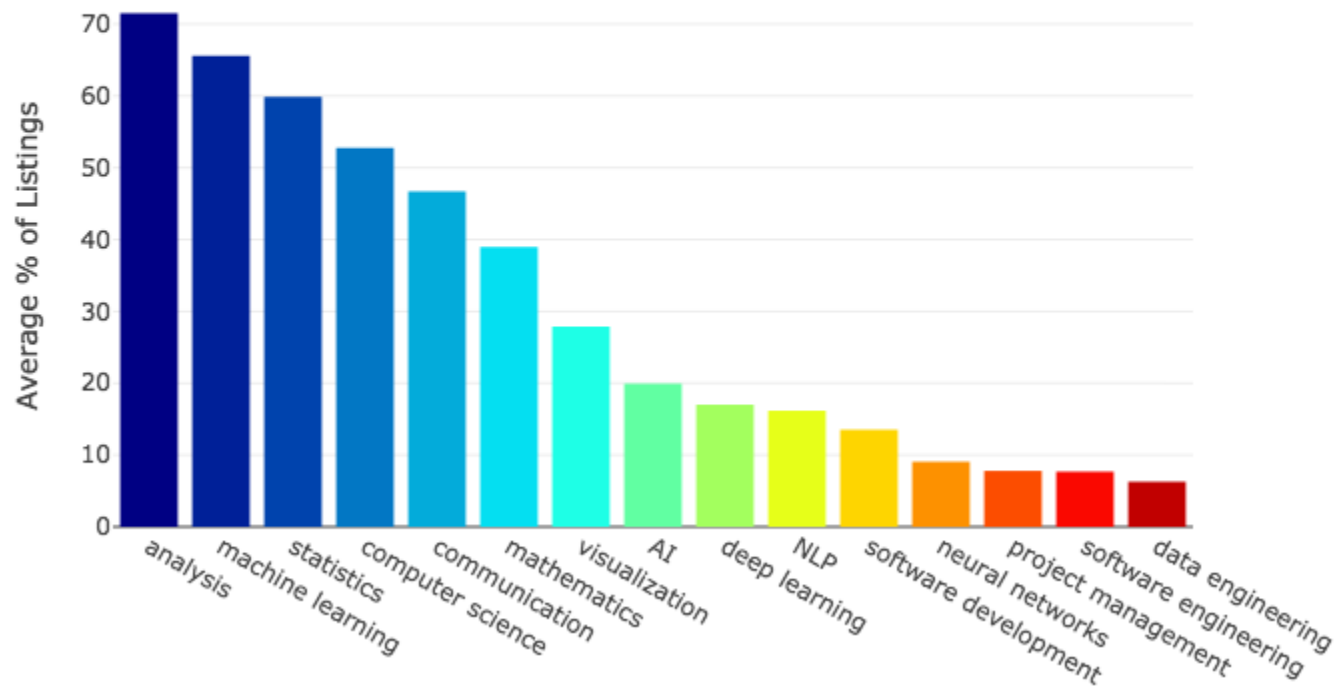
Población y muestra



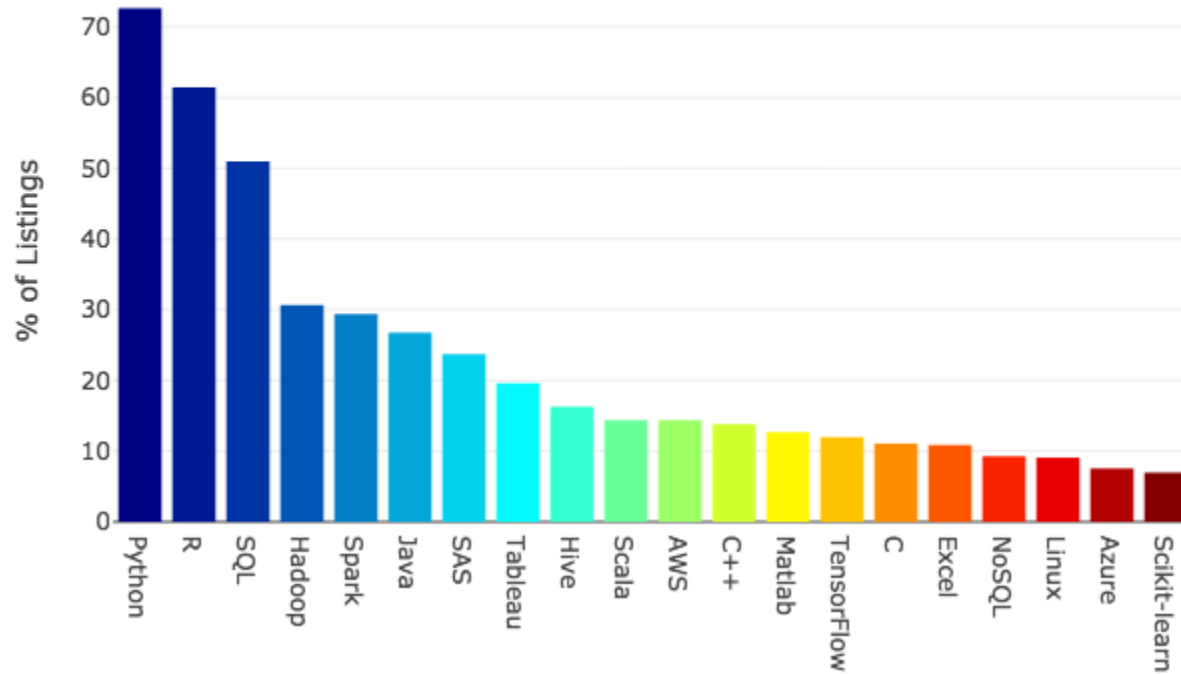
Ciencia de datos



General Skills in Data Scientist Job Listings



Top 20 Technology Skills in Data Scientist Job Listings



Modelos Mixtos

“Análisis de regresión lineal”

M.Sc. Henry Luis López García
Universidad Nacional Autónoma de Nicaragua, Managua
Facultad de Ciencia e Ingeniería



@Hen1985

Contenidos

- Modelo de regresión lineal
- Estimación de (β_0, β_1)
- Propiedad de los estimadores por mínimos cuadrados
- Estimación de (σ^2)
- Prueba de significación del modelo de regresión
- Diagnósticos de los residuos
- Coeficiente de correlación
- Estimación de (ρ)
- Prueba de significancia del coeficiente de correlación muestral (ρ)

Regresión lineal

La regresión es una técnica estadística para investigar y modelar la relación entre variables.

Propósito de la regresión lineal:

- Describir la regresión lineal entre y & x
- Determinar cuanta variación en y puede ser explicada por la relación con x
- Predecir valores nuevos de y usando nuevos valores de x

Regresión lineal

Consideremos el modelo de regresión lineal simple un modelo con solo un regresor x que tiene una relación con una respuesta y , donde la relación es una línea recta.

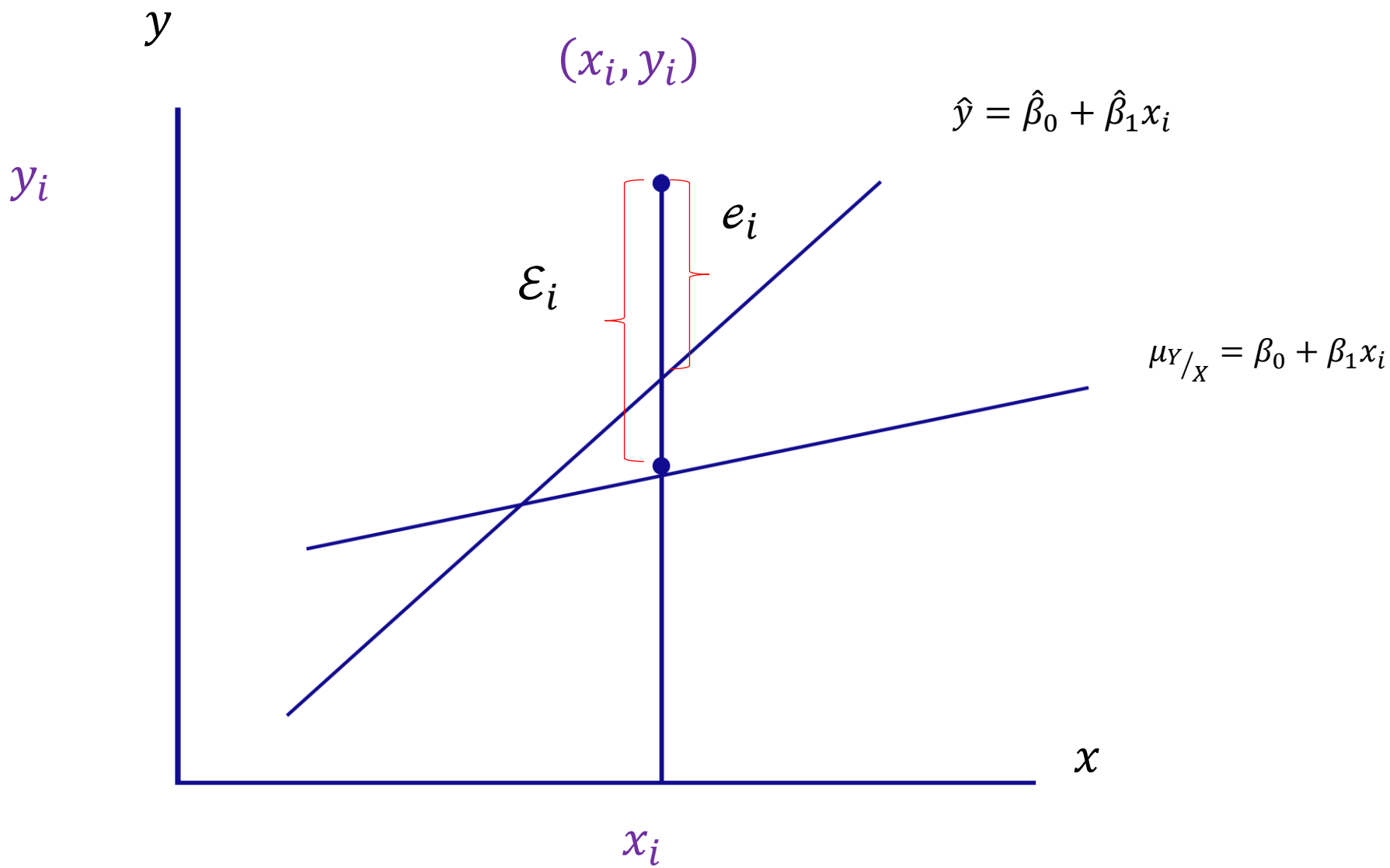
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (\text{ecuación 1})$$

Por tanto observando el modelo:

- Necesitamos estimar dos parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$.
- $\hat{\beta}_0$ es el intercepto, la media de la distribución de probabilidad de y cuando x es 0.

Regresión lineal

- $\hat{\beta}_1$ es a menudo llamado la pendiente, mide la tasa de cambio en y por una unidad de cambio en x .
- La estimación de los parámetros es a través de los mínimos cuadrados, lo que resuelve este modelo por medio minimizar e_i realmente $\sum_{i=1}^n e^2_i$.



Estimación de $(\beta_0 \text{ y } \beta_1)$

- Los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$ son desconocidos y se deben de estimar con los datos de la muestra, ahora supongamos que hay n pares de datos: $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$.
- Entonces para estimar $\hat{\beta}_0$ y $\hat{\beta}_1$ se usa el método de mínimo cuadrados, esto es, estimar $\hat{\beta}_0$ y $\hat{\beta}_1$ tales que la suma de los cuadrados de las diferencias entre las observaciones y_i y la línea recta sea mínima, según la ecuación puede escribirse

Estimación de $(\beta_0 \text{ y } \beta_1)$

- Considerando la $(y = \beta_0 + \beta_1 x + \varepsilon)$ es un modelo de regresión poblacional mientras que la ecuación 2, es un modelo muestral de regresión, escrito en términos de los n pares de datos (y_i, x_i) $i = 1, 2, \dots, n$

$$s(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \qquad \sum_{i=1}^n e^2_i$$

- Los estimadores por mínimos cuadrados, de β_0 y β_1 , que se designarán por $\hat{\beta}_0$ y $\hat{\beta}_1$, deben de satisfacer

Estimación de $(\beta_0 \text{ y } \beta_1)$

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Se simplifican estas dos ecuaciones se obtiene

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

Ecuaciones normales de mínimos cuadrados

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (\text{ecuación 3})$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad (\text{ecuación 4})$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Ecuaciones normales por mínimos cuadrados

Una forma cómoda de escribir la ecuación 4

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Suma corregida de los productos cruzados de las x_i & y_i

$$S_{xy} = \sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n} = \sum_{i=1}^n y_i (x_i - \bar{x})$$

Suma corregida de cuadrados de las x_i

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2$$

Residual (e_i)

- La diferencia entre el valor observado y_i & el valor ajustado correspondiente \hat{y}_i se le llama residual, matemáticamente, el i – *ésimo* residual es

$$e_i = y_i - \hat{y} = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, 2, \dots, n$$

(ecuación 5)

- Los residuales tienen un papel importante en la adecuación del modelo de regresión ajustado, y para detectar diferencias respecto a las hipótesis básicas.

Propiedad de los estimadores

- Los estimadores por mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ tienen algunas propiedades importantes, estas se describen:
- La suma de los residuales en cualquier modelo de regresión que contenga una ordenada al origen $\hat{\beta}_0$ siempre es igual a cero, esto es.

$$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0$$

Propiedad de los estimadores

- La suma de los valores observados y_i es igual a la suma de los valores ajustados \hat{y}_i

$$\sum_{i:1}^n y_i = \sum_{i:1}^n \hat{y}_i$$

- La línea de regresión de mínimos cuadrados siempre pasa por el **centroide** de los datos, que es el punto (\bar{y}, \bar{x}) .

Propiedad de los estimadores

- La suma de los residuales, ponderados por el valor correspondiente de la variable regresora, siempre es igual a cero:

$$\sum_{i=1}^n x_i e_i = 0$$

- La suma de los residuales, ponderados por el valor ajustado correspondiente, siempre es igual a cero:

$$\sum_{i=1}^n \hat{y}_i e_i = 0$$

Estimación de σ^2

- Además de estimar $\hat{\beta}_0$ y $\hat{\beta}_1$, se requiere un estimado de σ^2 para probar hipótesis y formar estimados de intervalos pertinentes al modelo de regresión, el estimado de σ^2 se obtiene de la **suma cuadrado residuales, o suma cuadrado del error**:

$$MSr_{Res} \sum_{i:1}^n e^2_i = \sum_{i:1}^n (y_i - \hat{y}_i)^2$$

- Se puede deducir una formula fácil de MSr_{Res} sustituyendo

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Estimación de σ^2

$$SS_{Res} = \sum_{i:1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}$$

$$SS_T = \sum_{i:1}^n y_i^2 - n\bar{y}^2 = \sum_{i:1}^n (y_i - \bar{y})^2$$

Es justo la suma de cuadrado corregida, de las observaciones de las respuestas por lo que:

$$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$$

Estimación de σ^2

- La suma de cuadrado de residuales tiene $n - 2$ grados de libertad, porque dos grados de libertad se asocian con los estimados $\hat{\beta}_0$ y $\hat{\beta}_1$ que se usan para obtener \hat{y}_i , por lo que un **estimador insesgado** de σ^2 es

$$MS_{Res} = \frac{SS_{Res}}{n - 2} = \hat{\sigma}^2$$

- La cantidad MS_{Res} se le llama **cuadrado medio residual** la raíz cuadrada de $\hat{\sigma}^2$, se llama el **error estándar de la regresión** y tiene la misma unidad que la variable de respuesta y .

Prueba de significancia de regresión

- Las siguientes hipótesis se relacionan con la significancia de la regresión, el no rechazar la $H_0: \beta_1 = 0$, implica que no hay relación lineal entre x & y . Esto puede implicar que x tiene muy poco valor para explicar la variación de y y que el mejor estimador x es $\hat{y} = \bar{y}$, o que la verdadera relación entre x & y no es lineal.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Prueba de significancia de regresión

- El procedimiento de prueba para $H_0: \beta_1 = 0$, se puede establecer, tan solo usando el estadístico t , en la siguiente ecuación:

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}, \text{ donde } se(\hat{\beta}_1) = \sqrt{\frac{MS_{Res}}{S_{xx}}}$$

- Rechazando $H_0: \beta_1 = 0$ si, $|t_0| > t_{\left(\frac{\alpha}{2}, n-2\right)}$

Prueba de significancia de regresión

- El procedimiento de prueba para $H_0: \beta_0 = 0$, se puede establecer, tan solo usando el estadístico t , en la siguiente ecuación:

$$t_0 = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)}, \text{ donde } se(\hat{\beta}_0) = \sqrt{MS_{Res} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

- Rechazando $H_0: \beta_0 = 0$ si, $|t_0| > t_{\left(\frac{\alpha}{2}, n-2\right)}$

Diagnósticos de los residuos

Las principales premisas que se han hecho hasta ahora al estudiar el análisis de regresión son las siguientes:

1. La relación entre la respuesta y y los regresores es lineal, al menos en forma aproximada.
2. El termino error ε tiene media cero.
3. El termino error ε tiene varianza σ^2 constante.
4. Los errores tienen distribución normal.

Diagnósticos de los residuos

Como se puede considerar que un residual es la desviación entre los datos y el ajuste, también es una medida de variabilidad de la variable de respuesta que no explica el modelo de regresión. También conviene imaginar que los residuales que los residuales son los valores realizados, u observados de los errores del modelo, por la que toda desviación de las premisas de los errores se debe reflejar en los residuos.

Diagnósticos de los residuos

- Residuales estandarizados, ya que la varianza aproximada de un residual se estima con MS_{Res} , el cuadrado medio de los residuales, un escalamiento lógico de los residuales sería el de los residuales estandarizados,

$$d_i = \frac{e_i}{\sqrt{MS_{Res}}}, \quad i = 1, 2, 3, \dots, n \quad MS_{Res} = \frac{\sum_{i=1}^n e_i^2}{n - p}$$

- Los residuales estandarizados tiene media cero y varianza aproximadamente unitaria, en consecuencia un residual estandarizado grande ($d_i > 3$).

Diagnósticos de los residuos

- **Residuales estudentizados**, Las violaciones de las premisas, del modelo, están con más probabilidad, en los puntos remotos, y pueden ser difíciles de detectar por inspecciones de los residuales ordinarios e_i por que en general sus residuales serán menores, entonces, un procedimiento lógico es examinar los residuales estudentizados,

$$r_i = \frac{e_i}{\sqrt{MS_{Res}(1-h_{ii})}}, \quad i = 1, 2, 3, \dots, n \quad h_{ii} = X(X^T X)^{-1} X^T$$

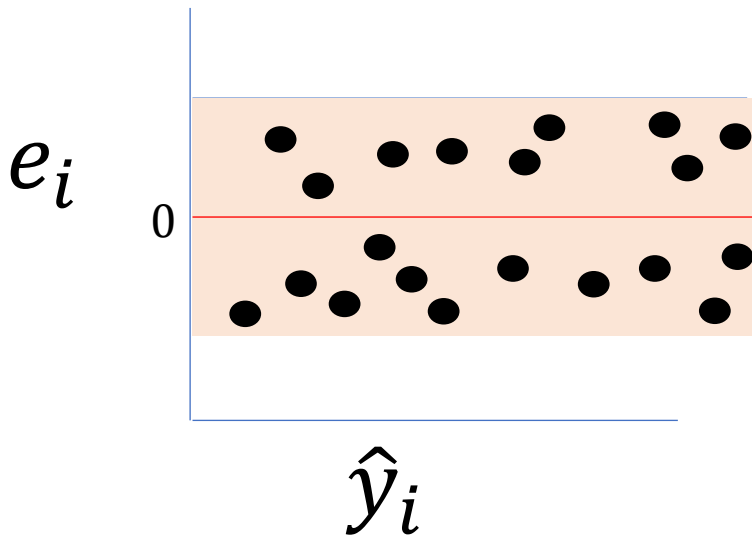
Agregar a X $V = (1, 1, 1, \dots, 1)$

Diagnósticos de los residuos

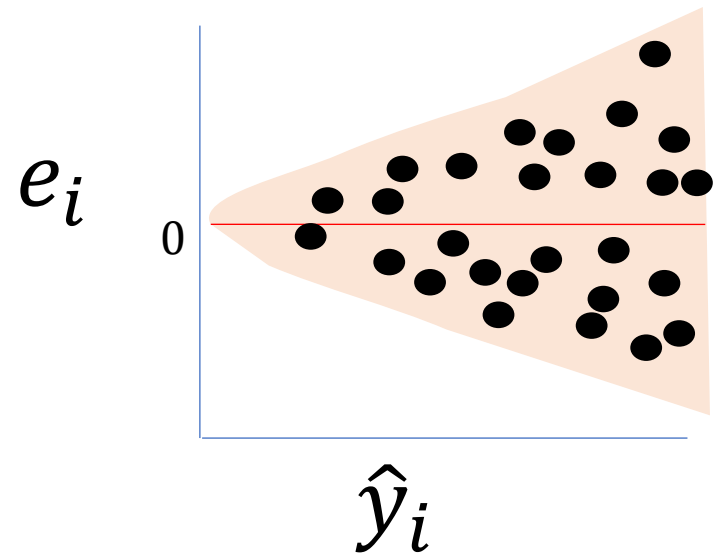
- **Residuales PRESS**, no es más que el residual ordinario ponderado por los elementos diagonal de la matriz de sombrero h_{ii} . Los residuales asociados con puntos para los h_{ii} es grande tendrán PRESS residuales grandes., estos serán por lo general puntos de gran influencia,

$$e_{(i)} = \frac{e_i}{1-h_{ii}}, \quad i = 1, 2, 3, \dots, n \qquad h_{ii} = X(X^T X)^{-1} X^T$$

Diagnósticos de los residuos

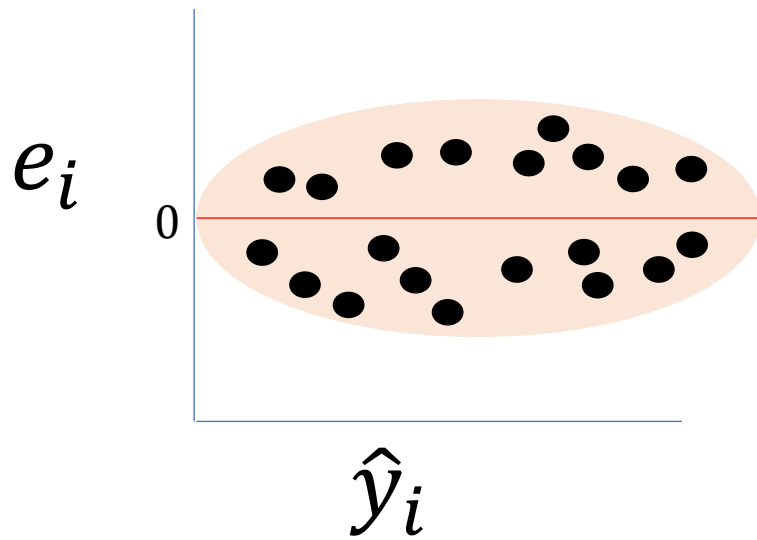


a)

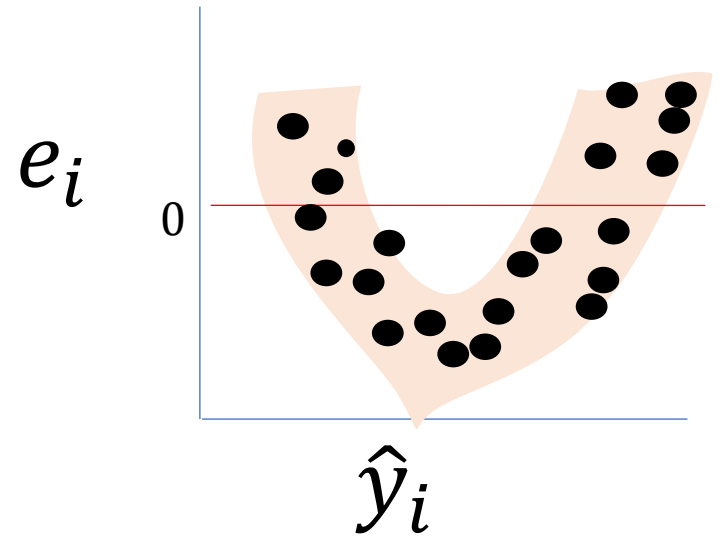


b)

Diagnósticos de los residuos



c)



d)

Modelos Mixtos

“Clase práctica # 1”

M.Sc. Henry Luis López García

Universidad Nacional Autónoma de Nicaragua, Managua

Facultad de Ciencia e Ingeniería



@Hen1985

Información de las variables

Data 2009			Data 2014		
Nombre	Etiqueta	Valores	Nombres	Etiqueta	Valores
S6PROD	Producto....		S6PROD	Producto....	
S6P9	Cada.....	1: Diario 2: Semanal 3: Quincenal 4: Mensual 5: Trimestral 6: Semestral 7: Anual 8:N/S 9: Ignorado	S7P4		1: Diario 2: Semanal 3: Quincenal 4: Mensual 5: Trimestral 6: Semestral 7: Anual 8:N/S 9: Ignorado
S6P5A	Que cantidad..		S7P5A	Que cantidad..	
S6P6	Cuánto.....		S7P6A	Cuánto....	

Procedimiento para construir gastosm

*EMNV09-06 PARTE A DE LA SECCION 6.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Calcular variable...

Calcular variable

Variable objetivo: gastosmm = Expresión numérica: S6P6*28

Tipo y etiqueta...

Numero de cuestionario...
Formulario [I00A]
Hogar [I00B]
Departamento [I01]
Municipio [I02]
Siete Regiones [domini...]
Regiones [dominio4]
Estrato [I05]
Área de Residencia [I06]
Resultado entrevista [R...]
Producto [S6PROD]
Algún miembro de este ...
Cada cuánto compra [S...]
Que cantidad de product...
Unidad de medida [S6P...]
Cuánto pagaron en total ...
En estos últ. 15 días obt...
Cada cuánto obtienen o ...
Que cantidad de product

Grupo de funciones:
Todo
Aritméticas
CDF y CDF no centrada
Conversión
Fecha/hora actual
Cálculo de fechas
Creación de fechas

Funciones y variables especiales:

Nueva variable

Cuánto pagaron en total por...

Si...

Aceptar Pegar Restablecer Cancelar Ayuda

Procedimiento

Calcular variable: Si los casos

☐ Incluir todos los casos

☒ Incluir si el caso satisface la condición:

S6P4 = 1 → 1: Mensual

→ Cada cuánto compra

Grupo de funciones:

- Todo
- Aritméticas
- CDF y CDF no centrada
- Conversión
- Fecha/hora actual
- Cálculo de fechas
- Creación de fechas
- Extracción de fechas
- GL inversos

Funciones y variables especiales:

Continuar Cancelar Ayuda

Lista de variables:

- Numero de cuestionario [I0...
- Formulario [I00A]
- Hogar [I00B]
- Departamento [I01]
- Municipio [I02]
- Siete Regiones [dominio7]
- Regiones [dominio4]
- Estrato [I05]
- Área de Residencia [I06]
- Resultado entrevista [RES...
- Producto [S6PROD]
- Algún miembro de este ho...
- Cada cuánto compra [S6P4]
- Que cantidad de producto ...
- Unidad de medida [S6P5B]
- Cuánto pagaron en total po...
- En estos últ. 15 días obtuv...
- Cada cuánto obtienen o le...
- Que cantidad de producto ...
- Unidad de medida [S6P9B]
- Cuánto tendría que pagar ...
- Unidad Primaria de Muestr...
- Factor de expansión con c...
- Pobre (todos) 2009 [pob.ge]
- Pobreza 2009 [pobreza]
- No pobre 2009 [pob.no]
- Gastos_m
- S6PROD = 28 (FILTER) [filt...
- gastosm

Procedimiento

Calcular variable

Variable objetivo: gastosmm = Expresión numérica: S6P6*28

Tipo y etiqueta...

Numero de cuestionario [I00]
Formulario [I00A]
Hogar [I00B]
Departamento [I01]
Municipio [I02]
Siete Regiones [dominio7]
Regiones [dominio4]
Estrato [I05]
Área de Residencia [I06]
Resultado entrevista [RESULTA...]
Producto [S6PROD]
Algún miembro de este hogar c...
Cada cuánto compra [S6P4]
Que cantidad de producto comp...
Unidad de medida [S6P5B]
Cuánto pagaron en total por [S6...
En estos últ. 15 días obtuvo pro...
Cada cuánto obtienen o les dan...
Que cantidad de producto obtuvi

+ < > 7 8 9
- <= >= 4 5 6
* = ~= 1 2 3
/ & | 0 .
** ~ () Suprimir

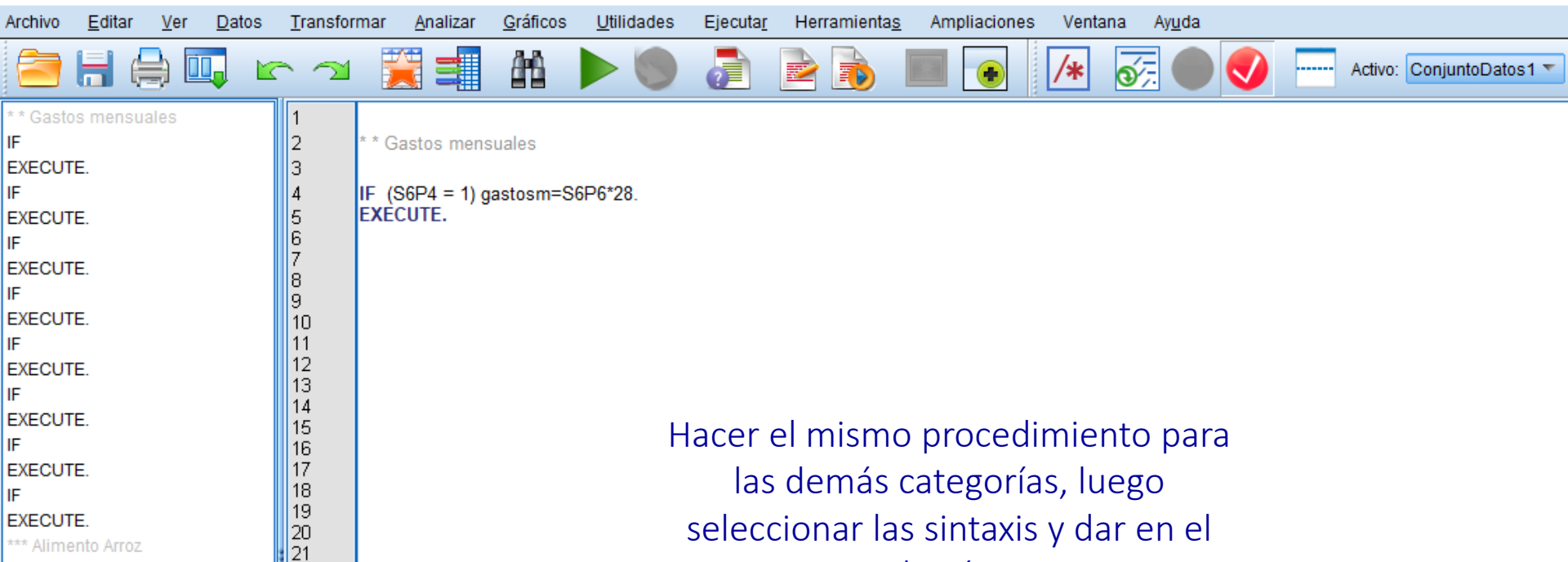
Grupo de funciones:
Todo
Aritméticas
CDF y CDF no centrada
Conversión
Fecha/hora actual
Cálculo de fechas
Creación de fechas

Funciones y variables especiales:

Si... S6P4 = 1

Aceptar Pegar Restablecer Cancelar Ayuda

Procedimiento



Hacer el mismo procedimiento para
las demás categorías, luego
seleccionar las sintaxis y dar en el
botón



Construir la variable `gastosc` data 2009 y 2014

IF (`S6P4` = 1) `gastosc`=`S6P6`*28.

EXECUTE.

IF (`S6P4` = 2) `gastosc`=`S6P6`*4.

EXECUTE.

IF (`S6P4` = 3) `gastosc`=`S6P6`*2.

EXECUTE.

IF (`S6P4` = 4) `gastosc`=`S6P6`*1.

EXECUTE.

IF (`S6P4` = 5) `gastosc`=`S6P6`/3.

EXECUTE.

IF (`S6P4` = 6) `gastosc`=`S6P6`/6.

EXECUTE.

IF (`S6P4` = 7) `gastosc`=`S6P6`/12.

EXECUTE.

Verificar el archivo de trabajo para la variable

- S6PROD: **Producto** (emvn-2009)
- S6PROD: **Producto** (emvn-2014)

Solicitar el gasto promedio mensual para los alimentos de la canasta básica

*EMNV09-06 PARTE A DE LA SECCION 6.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ayuda

Definir propiedades de variables...
Definir nivel de medición para desconocido...
Copiar propiedades de datos...
Nuevo atributo personalizado...
Definir fecha y hora...
Definir conjuntos de respuestas múltiples...
Validación
Identificar casos duplicados...
Identificar casos atípicos...
Comparar conjuntos de datos...
Ordenar casos...
Ordenar variables...
Transponer...
Ajustar el ancho de las cadenas de distintos archivos
Fusionar archivos
Reestructurar...
Ponderaciones Rake...
Coincidencia de puntuación de propensión...
Coincidencia de control de casos...
Agregar...
Diseño ortogonal
Dividir en archivos
Copiar conjunto de datos
Segmentar archivo...
Seleccionar casos...
Ponderar casos...

Nombre
1 I00
2 I00A
3 I00B
4 I01
5 I02
6 dominio7
7 dominio4
8 I05
9 I06
10 RESULTAD
11 S6PROD
12 S6P3
13 S6P4
14 S6P5A
15 S6P5B
16 S6P6
17 S6P7
18 S6P8
19 S6P9A
20 S6P9B
21 S6P10
22 upm
23 Peso2
24 pob.ge
25 pobreza
26 pob.no

Seleccionar casos

Seleccionar

- ☒ Todos los casos
- ☐ Si se satisface la condición
Si...
- ☐ Muestra aleatoria de casos
Ejemplo...
- ☐ Basándose en el rango del tiempo o de los casos
Rango...
- ☐ Usar variable de filtro:
[]

Resultado

- ☒ Descartar casos no seleccionados
- ☐ Copiar casos seleccionados a un nuevo conjunto de datos
Nombre de conjunto de datos: []
- ☐ Eliminar casos no seleccionados

Estado actual: No filtrar casos

Aceptar Pegar Restablecer Cancelar Ayuda



Seleccionar casos: Si la opción



S6PROD = 1

Grupo de funciones:

Todo

Aritméticas



Seleccionar casos

Seleccionar

☐ Todos los casos

☒ Si se satisface la condición

Si...

S6PROD = 1

☐ Muestra aleatoria de casos

Ejemplo...

☐ Basándose en el rango del tiempo o de los casos

Rango...

☐ Usar variable de filtro:



Resultado

☒ Descartar casos no seleccionados

☐ Copiar casos seleccionados a un nuevo conjunto de datos

Nombre de conjunto de datos:

☐ Eliminar casos no seleccionados

Estado actual: No filtrar casos

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Archivo Editar Ver Datos Transformar **Analizar** Gráficos Utilidades Ampliaciones Ventana Ayuda

Informes

Estadísticos descriptivos

Estadísticas Bayesianas

Tablas

Comparar medias

Modelo lineal general

Modelos lineales generalizados

Frecuencias...

Descriptivos...

Explorar...

Tablas cruzadas...

Análisis TURF

Razón...

Gráficos P-P...

	Nombre	Tipo	Anchor
1	I00	Número	8
2	I00A	Número	5
3	I00B	Número	1
4	I01	Número	2

Frecuencias

Variables:

gastosm

Estadísticos...

Gráficos...

Formato...

Estilo...

Simular muestreo...

Mostrar tablas de frecuencias

Aceptar Pegar Restablecer Cancelar Ayuda

Frecuencias: Estadísticos

Valores percentiles

☐ Cuantiles

☐ Puntos de corte para: 10 grupos iguales

☐ Percentiles:

Añadir

Cambiar

Eliminar

Tendencia central

☒ Media

☐ Mediana

☐ Moda

☐ Suma

☐ Los valores son puntos medios

Dispersión

☐ Desviación estándar

☐ Varianza

☐ Rango

☐ Mínimo

☐ Máximo

☐ Error estándar media

Caracterizar distribución posterior

☐ Asimetría

☐ Curtosis

Continuar Cancelar Ayuda

```

7 IF (S6P4 = 1) gastosm=S6P6*28.
8 EXECUTE.
9
10 IF (S6P4 = 2) gastosm=S6P6*4.
11 EXECUTE.
12
13 IF (S6P4 = 3) gastosm=S6P6*2.
14 EXECUTE.
15
16 IF (S6P4 = 4) gastosm=S6P6*1.
17 EXECUTE.
18
19 IF (S6P4 = 5) gastosm=S6P6/3.
20 EXECUTE.
21
22 IF (S6P4 = 6) gastosm=S6P6/6.
23 EXECUTE.
24
25 IF (S6P4 = 7) gastosm=S6P6/12.
26 EXECUTE.
27
28 *** Alimento Arroz
29
30
31
32 USE ALL.
33 COMPUTE filter_$(S6PROD = 1).
34 VARIABLE LABELS filter_$ 'S6PROD = 1 (FILTER)'.
35 VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
36 FORMATS filter_$ (f1.0).
37 FILTER BY filter_$.
38 EXECUTE.
39
40 ►
41 FREQUENCIES VARIABLES=gastosm
42 /FORMAT=NOTABLE
43 /ORDER=ANALYSIS.
44

```



Solicitar el gasto promedio mensual para los alimentos de la canasta básica

Ejemplo

** Arroz en grano

USE ALL.

COMPUTE filter_\$=(S6PROD = 9.0).

VARIABLE LABELS filter_\$ 'S6PROD = 9.0 (FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_\$ (f1.0).
FILTER BY filter_\$.
EXECUTE.

FREQUENCIES VARIABLES=gastosm

/FORMAT=NOTABLE

/STATISTICS=VARIANCE MEAN

/ORDER=ANALYSIS.

Modelos Mixtos

“R: Introducción”

M.Sc. Henry Luis López García
Universidad Nacional Autónoma de Nicaragua, Managua
Facultad de Ciencia e Ingeniería



@Hen1985

Qué es R?

- Leguaje para la computación estadística
- Variedad de técnicas estadísticas
- Similar al entorno de lenguaje S
- R fue escrito inicialmente por Robert Gentleman y Ross Ihaka, también conocido como " R & R " del Departamento de Estadísticas de la Universidad de Auckland. Desde mediados de 1997.
- Capacidad de visualización
- Altamente extensible

Ventajas?

- Código abierto! ¡Gratis!
- Master en gráficos
- Interfaz de línea de comando
- Reproducibilidad a través de scripts R
- Extensivo
- Comunidad (contribuciones de todo el mundo)

Desventajas?

- Fácil de aprender, difícil de dominar
- Interfaz de línea de comandos desalentadora al principio
- Código mal escrito difícil de leer / mantener

Descarga

R

- <https://www.r-project.org/>

RStudio

- <https://rstudio.com/>

Modelos Mixtos

“Análisis de regresión lineal”

M.Sc. Henry Luis López García
Universidad Nacional Autónoma de Nicaragua, Managua
Facultad de Ciencia e Ingeniería



@Hen1985