

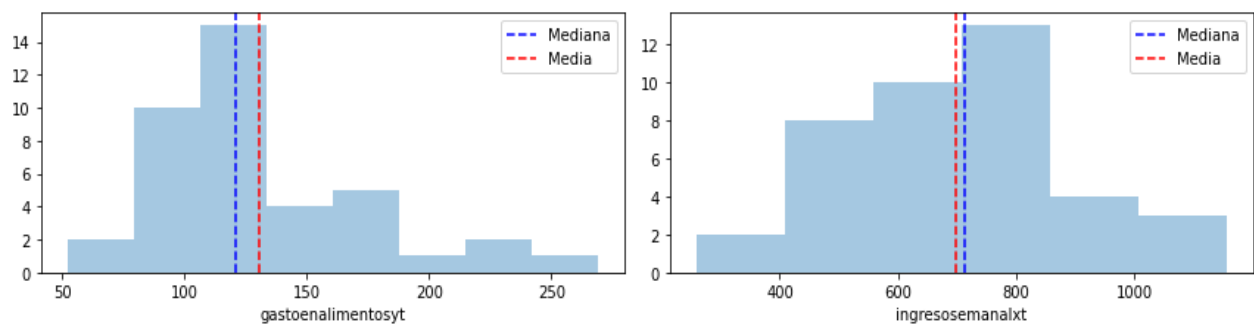
# TP N° 1 – Econometría

1 y 2)

Verificación de datos faltantes en el dataset:

```
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   gastoalimentosyt      40 non-null    float64
1   ingreso semanalxt      40 non-null    float64
dtypes: float64(2)
memory usage: 768.0 bytes
```

En el dataset hay 40 observaciones y 2 columnas sin ningún valor nulo en las variables X e Y.



	min	q25	q50	q75	max	mean	range	std	skew	kurt
<b>gastoalimentosyt</b>	52.25000	100.365000	120.70500	147.685000	269.03	130.313000	216.78000	45.158568	1.075598	1.451209
<b>ingreso semanalxt</b>	258.29999	560.624987	712.29999	820.274995	1154.60	697.999997	896.30001	198.226948	0.227275	0.150349

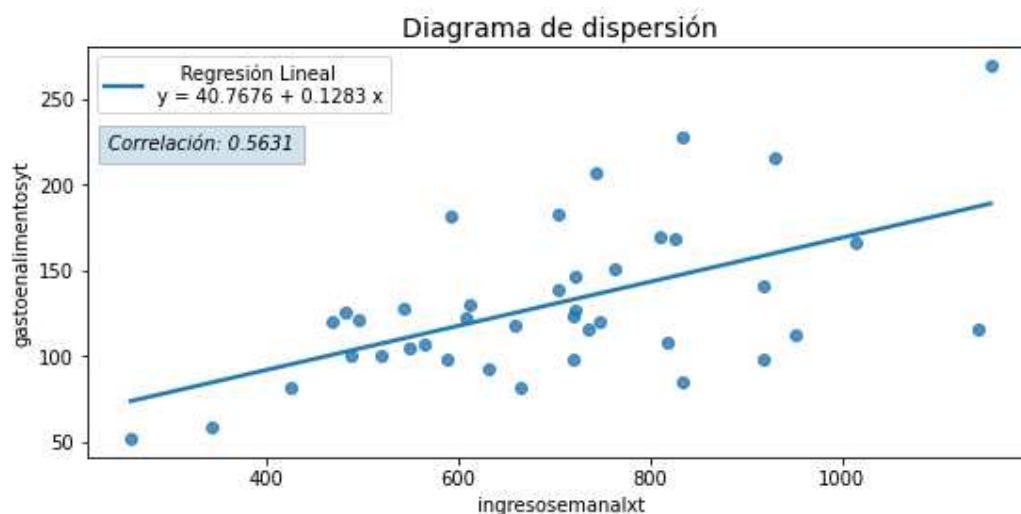
En las medidas de tendencia central (media y mediana/q50) se puede observar que la media y mediana de la variable dependiente se encuentran separadas siendo la media > mediana y presentando una diferencia de 148.325 y 68.455, valores calculados como (máximo - mediana) y (mediana - mínimo) respectivamente, esto significa que el valor máximo de las observaciones se encuentra un poco más que el doble de lejos con respecto a la mediana en comparación al valor mínimo. Éstos son indicios de asimetría positiva. En cambio en la variable independiente las medidas de tendencia central se encuentran casi juntas, como debería ser en una distribución normal. La diferencia entre el valor máximo-mediana y mediana-mínimo es de 442.3 y 454.0 respectivamente, indicando una distribución bastante simétrica.

En cuanto a las medidas de dispersión (rango y desviación estándar), las mismas son mayores en la variable exógena que en la endógena.

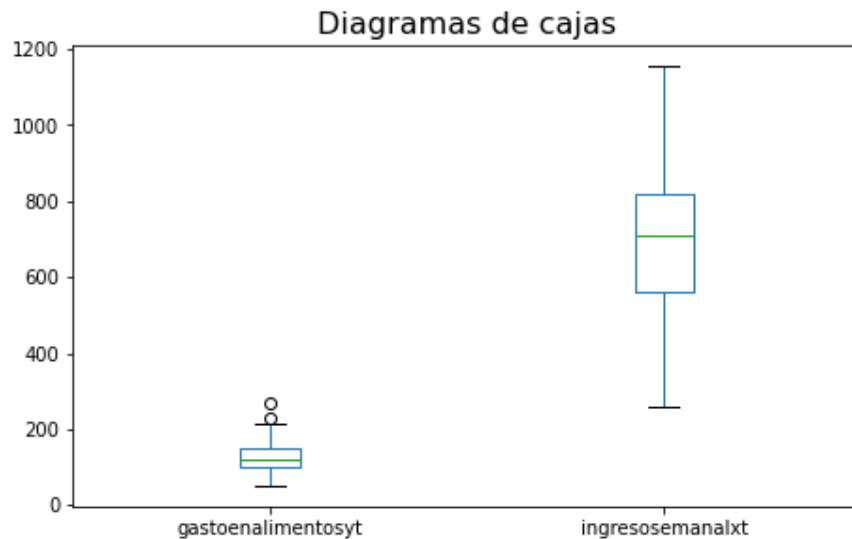
El histograma de 'gastosenalimentosyt' indica que las observaciones de ésta variable tienen bastante asimetría positiva con un valor de 1.075598. La kurtosis tiene un valor de 1.451209 indicando que la distribución es bastante leptocurtica, distribución muy estrecha con la mayoría de las observaciones concentradas alrededor de la media.

Por otro lado, el histograma de 'ingresosemanalxt' los datos parecen tener una distribución que se aproxima a la normal, bastante simétrica. El valor de la asimetría es de 0.227275, casi igual al de un conjunto de datos perfectamente simétricos (asimetría = 0). La kurtosis tiene un valor de 0.150349 indicando que la distribución de los datos se aproxima a la normal (distribución mesocúrtica).

3)



En el gráfico se puede observar que a medida que aumentan los ingresos semanales los consumidores tienden a gastar más plata en alimentos. La correlación de Pearson entre ellas es de 0.5631, reafirmando que hay correlación moderada positiva. Hay una relación lineal entre las variables en cuestión, dicho punto es importante por el supuesto de linealidad entre los datos. Otra observación es que a medida que aumentan los ingresos semanales aumenta la dispersión o rango vertical en los gastos en alimentos, dando lugar a posibles outliers y a que el supuesto de normalidad en la distribución de los datos  $p(y|x)$  no se esté cumpliendo. Los outliers los podemos verificar con unos diagramas de caja.



Efectivamente podemos observar outliers en la variable dependiente (gastos en alimentos). Estas observaciones anormales pueden distorsionar los coeficientes de regresión y limitar la capacidad del modelo para predecir correctamente los casos normales/habituales.

4)

```

=====
                        OLS Regression Results
=====
Dep. Variable:          gastoenalimentosyt    R-squared:                0.317
Model:                  OLS                  Adj. R-squared:           0.299
Method:                 Least Squares        F-statistic:             17.65
Date:                  Tue, 02 Nov 2021      Prob (F-statistic):      0.000155
Time:                  22:24:46              Log-Likelihood:          -201.03
No. Observations:      40                   AIC:                    406.1
Df Residuals:          38                   BIC:                    409.4
Df Model:              1
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	40.7676	22.139	1.841	0.073	-4.050	85.585
ingresosemanalxt	0.1283	0.031	4.201	0.000	0.066	0.190

```

=====
Omnibus:                 1.365    Durbin-Watson:           2.370
Prob(Omnibus):           0.505    Jarque-Bera (JB):        1.077
Skew:                    0.397    Prob(JB):                0.584
Kurtosis:                2.874    Cond. No.                2.68e+03
=====

```

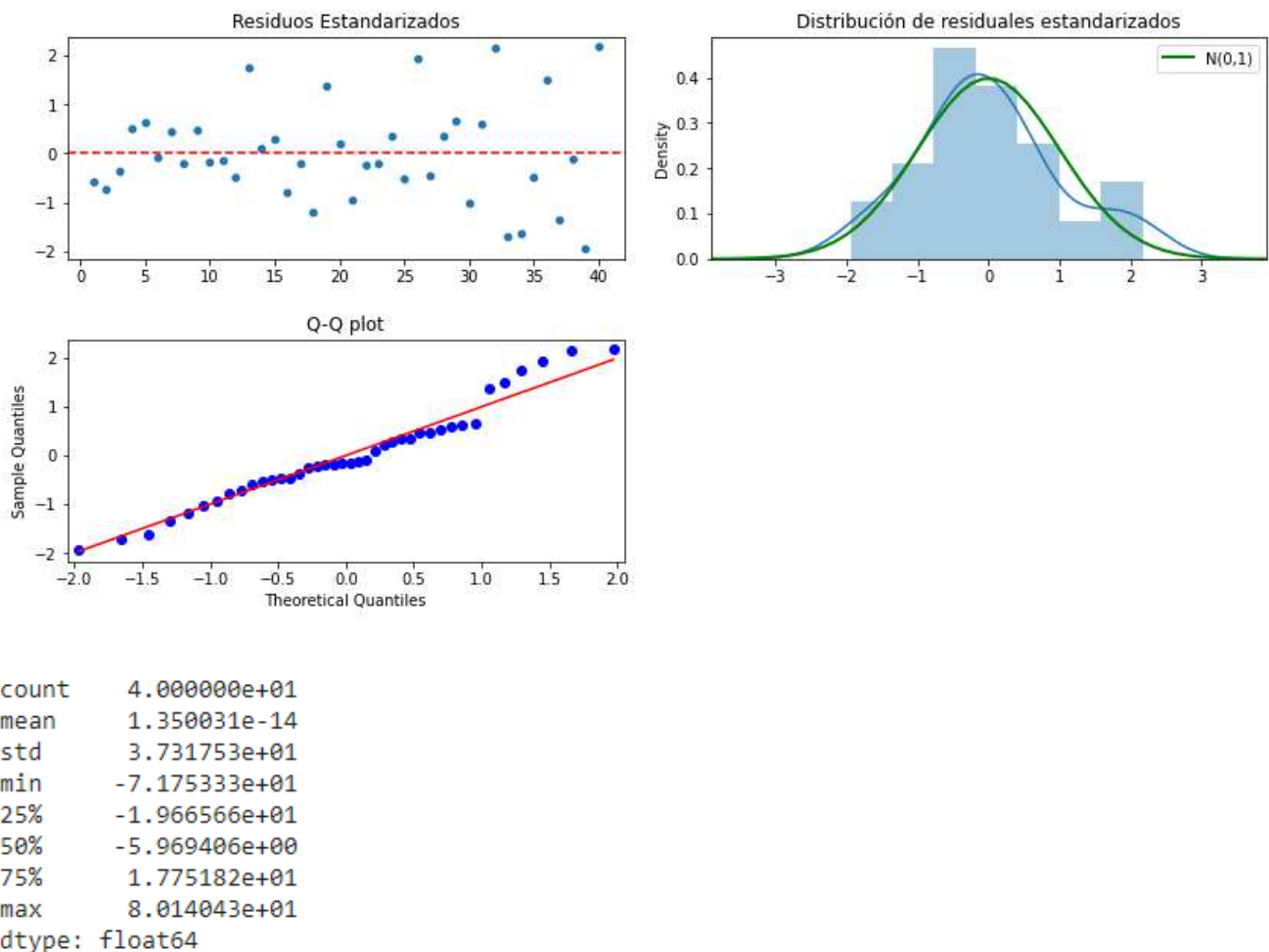
Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
 [2] The condition number is large, 2.68e+03. This might indicate that there are strong multicollinearity or other numerical problems.

La dirección entre ambas variables es positiva, es decir cuando incrementa en una unidad la variable independiente 'ingresosemanalxt' la variable dependiente 'gastoenalimentosyt' aumenta 0.1283.

Cuando la variable predictora incluida en el modelo es igual a 0, el valor de la variable dependiente es igual a 40.7676.

5)



En el gráfico de \*Q-plot\* y el de \*Distribución de residuales estandarizados\* pareciera que la distribución de los residuales se aproxima a una normal, aun así podemos ver en ellos que hay asimetría positiva. En el Q-plot los residuales, aproximadamente a partir del cuartil 1, se desvían por encima de la distribución normal teórica (línea roja); mientras que en el histograma hay residuos que se sitúan por encima de +2 desviaciones estándar llegando a superar el 95.4% de los demás residuos.

En el gráfico de los \*Residuos Estandarizados\* vemos que para cada observación de la muestra aumenta la varianza de los residuos a medida que aumentan los valores en X, presentando aparentemente heterocedasticidad.

6)

### Prueba de heterocedasticidad - Gleiser test

$H_0$  = La homocedasticidad está presente en el modelo de regresión

$H_a$  = La heteroscedasticidad está presente en el modelo de regresión

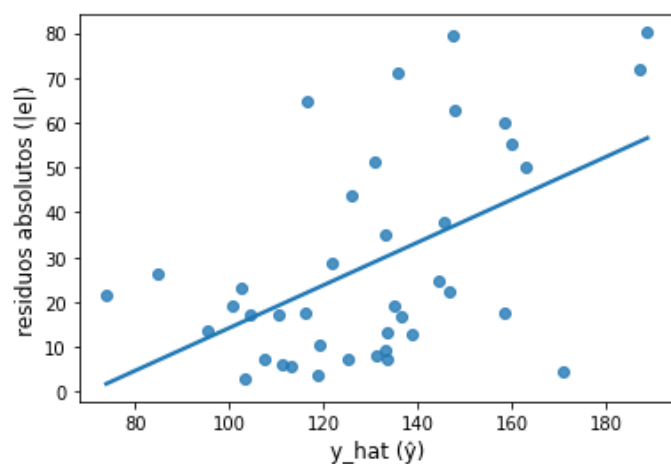
```

OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.267
Model:                  OLS    Adj. R-squared:       0.248
Method:                 Least Squares    F-statistic:       13.84
Date:                   Tue, 02 Nov 2021    Prob (F-statistic): 0.000642
Time:                   22:24:46    Log-Likelihood:    -176.26
No. Observations:       40    AIC:              356.5
Df Residuals:           38    BIC:              359.9
Df Model:                1
Covariance Type:        nonrobust
=====
               coef    std err          t      P>|t|      [0.025    0.975]
-----
const         -33.4680     17.006     -1.968     0.056    -67.895     0.959
0              0.4767      0.128      3.720     0.001      0.217     0.736
=====
Omnibus:                 1.231    Durbin-Watson:       2.025
Prob(Omnibus):            0.540    Jarque-Bera (JB):     1.219
Skew:                     0.372    Prob(JB):             0.544
Kurtosis:                 2.580    Cond. No.             701.
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



Luego de estimar una regresión entre los 'valores absolutos de los residuales' y los 'Y sombrero', hacemos una prueba de significancia sobre el coeficiente de la pendiente. Dado que  $p < 0.05$  rechazamos  $H_0$ , esto significa que la tendencia observada no se explica por casualidad bajo un modelo homocedástico.

#### Prueba de heterocedasticidad - Breush Pagan

$H_0$  = La homocedasticidad está presente en el modelo de regresión

$H_a$  = La heteroscedasticidad está presente en el modelo de regresión

Luego de correr la prueba de hipótesis el valor de 'p' nos da 0.00052. Dado que  $p < 0.05$  rechazamos la hipótesis nula a un nivel de significancia de 5%, por lo tanto asumimos que los residuos no se distribuyen con varianza constante. Hay heteroscedasticidad en el modelo de regresión.

#### 7)

\* R-squared: el 31.7% de la varianza de `gastoenalimentosyt` es capturada por `ingresosemanalxt`.

\* Prob (F-statistic): Al ser  $p(F) < 0.05$  rechazamos la hipótesis nula de que todos los coeficientes de la regresión (menos la constante) sean igual a 0, es decir que ningún beta esté relacionado con Y.

\* std err: Al ser el error standard 0.031, si con muestras aleatorias de igual tamaño se construyen 100 intervalos, se espera que 95 de ellos contengan la verdadera pendiente de `ingresosemanalxt` entre  $[0.1283 - 2.0227 * 0.031 < \text{Beta} < 0.1283 + 2.0227 * 0.031]$  (intervalo de confianza), lo que indica que hay una relación positiva entre las variables en estudio.

\*  $P > |t|$ : al ser  $p < 0.05$  para el predictor '`ingresosemanalxt`', rechazamos la hipótesis nula de que el coeficiente sea igual a 0. Es decir, que es estadísticamente significativo éste parámetro en nuestro modelo de regresión.

\* Intervalo de confianza: Con un nivel de confianza del 95.0%, si se construyen 100 intervalos con muestras aleatorias de igual tamaño, se espera que 95 de ellos contengan el rango razonable para nuestro parámetro de '`ingresosemanalxt`' es entre 0.066 y 0.190, el valor 0 para nuestro beta no se encuentra dentro del intervalo de confianza.

#### 8)

Observaciones	gastoenalimentosyt	y_sombrero	residuos
1	52,25	73,9045	-21,6545
2	58,32	84,78338	-26,4634

3	81,79	95,29021	-13,5002
4	119,9	100,7425	19,15752
5	125,8	102,7181	23,08188
6	100,46	103,3339	-2,87391
7	121,51	104,4628	17,04715
8	100,08	107,4007	-7,32066
9	127,75	110,4668	17,28325
10	104,94	111,1595	-6,21951
11	107,48	113,1993	-5,7193
12	98,48	116,2397	-17,7597
13	181,21	116,6246	64,5854
14	122,23	118,6772	3,552777
15	129,57	119,1776	10,39246
16	92,84	121,7177	-28,8777
17	117,92	125,3867	-7,46672
18	82,13	125,9512	-43,8212
19	182,28	131,1084	51,17161
20	139,13	131,1854	7,944639
21	98,14	133,1097	-34,9697
22	123,94	133,1353	-9,19535
23	126,31	133,4304	-7,12041
24	146,47	133,4304	13,03959
25	115,98	134,9827	-19,0027
26	207,23	136,0218	71,20816

27	119,8	136,6889	-16,8889
28	151,33	138,6902	12,63976
29	169,51	144,707	24,80301
30	108,03	145,7718	-37,7418
31	168,9	146,6826	22,21737
32	227,11	147,6704	79,43955
33	84,94	147,7602	-62,8202
34	98,7	158,5493	-59,8493
35	141,06	158,5493	-17,4893
36	215,4	160,0246	55,37535
37	112,89	162,8598	-49,9698
38	166,25	170,8522	-4,6022
39	115,43	187,1833	-71,7533
40	269,03	188,8896	80,14043

La suma de los residuos del modelo es: 5.684341886080801e-13. El valor se aproxima a 0.

**9)**

La elasticidad ingreso de la demanda es: 0.6872%. A partir de éste valor podemos deducir que estamos ante un bien normal de primera necesidad ya que el valor obtenido es positivo y no supera la unidad.

Éste valor indica que ante un incremento del 1% en los ingresos semanales, el gasto en alimentos aumenta un 0.6872 %

**10)**

$$y = 40.76755999820226 + 0.12828859600524858 * 750$$

$$y = 136.984$$