

TRABAJO EN CLASE : MÉTODOS LINEALES

Presentado por: DANNY ALEXANDER RAMIREZ ROJAS

Profesora: EDDY JOHANA FAJARDO

Fecha: 19/09/2024

Actividad TAREA 1

Teniendo en cuenta la base de datos dada, realizar:

A Diagrama de dispersión utilizando RStudio (señalar en el gráfico tanto la media de la variable dependiente como la media de la variable regresora). Explicar si existe relación lineal entre las variables.

B

Si existe relación lineal entre las variables, explicar si la relación es directa o inversa.

C Estimar el valor de la covarianza entre las dos variables. Interpretar este valor

D A través de métodos gráficos Histograma, Caja y bigotes, qqplot tratar de inferir normalidad.

E A través de métodos analíticos encontrar Normalidad , definir el método utilizado y hallar los siguientes valores: Estadístico de prueba, Nivel de Significancia y Valor P

F Analizar la significancia de la relación Gráficamente y luego comprándola analíticamente con el método determinado según su Normalidad , definiendo su valores P Y r

Vamos a analizar las siguientes correlaciones de la base de datos del ICFES para el Área metropolitana de Bucaramanga

ÍTEM CORREL	Variable Regresora o Independiente	Variable Dependiente
1	ESTU_INSE_INDIVIDUAL	PUNT_GLOBAL
2	ESTU_INSE_INDIVIDUAL	PUNT_INGLES
3	ESTU_INSE_INDIVIDUAL	PUNT_C_NATURALES
4	ESTU_INSE_INDIVIDUAL	PUNT_MATEMATICAS
5	ESTU_INSE_INDIVIDUAL	PUNT_LECTURA_CRITICA

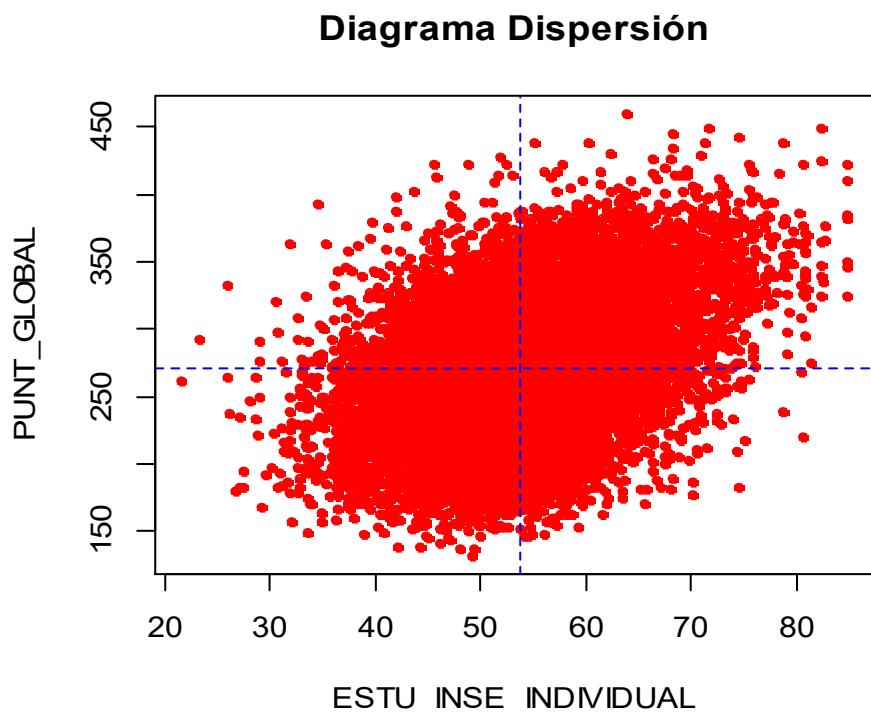
1

Variable Regresora o Independiente Variable Dependiente

ESTU_INSE_INDIVIDUAL

PUNT_GLOBAL

A.1



B.1

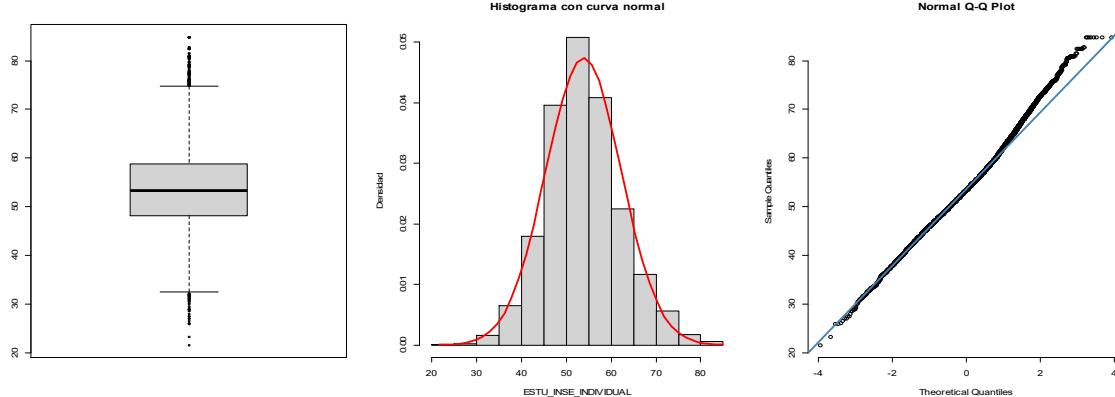
Al parecer hay una relación lineal entre las variables aunque débil.

C.1

	ESTU_INSE_INDIVIDUAL	PUNT_GLOBAL
mean	53,85	271,2
n =	12374	

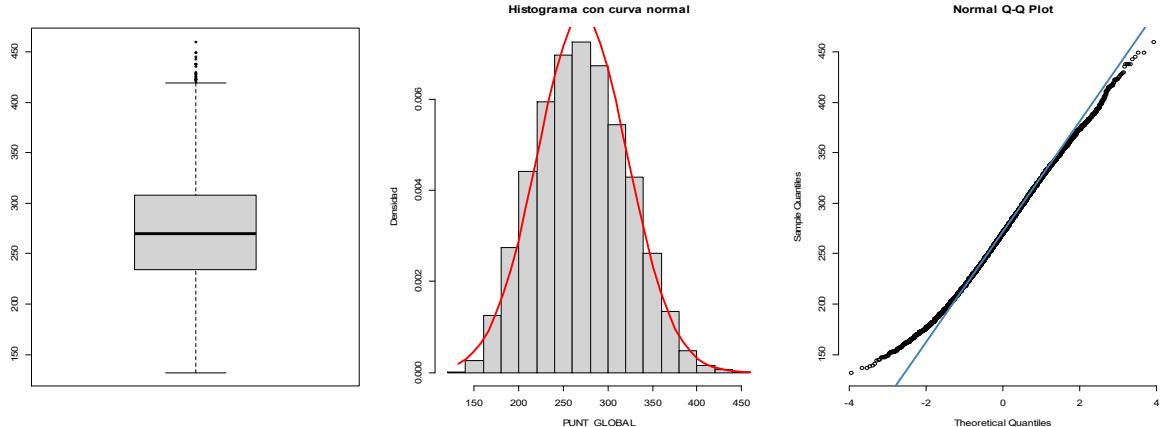
Lo que nos indica la covarianza es que la relación es Directa.

D.1 - Validar Normalidad gráficamente de la ESTU_INSE_INDIVIDUAL Variable :



Gráficamente podemos apreciar que esta variable tiene un comportamiento aproximadamente normal en los tres gráficos, aunque en el qqplot se ve en la cola derecha los valores se están alejando de la media

- Validar Normalidad gráficamente de la Variable : PUNT_GLOBAL



Gráficamente podemos apreciar que esta variable tiene un comportamiento aproximadamente normal en los tres gráficos, aunque en el qqplot se ve en la cola derecha e izquierda los valores se están alejando de la media.

- E.1 - Validar Normalidad numéricamente de la ESTU_INSE_INDIVIDUAL Variable :

H_0 = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

n = 12374 $n > 50$ por lo tanto el método a utilizar es (KS)

Método a utilizar : KOLMOGOROV - SMIRNOV (KS)

RESULTADOS	Est. De Prueba (D)	P-Valor
	0,0344	3,466E-13

Podemos apreciar un P-Valor muy por debajo del Nivel de Significancia ($\alpha = 0,05$) por lo que Rechazamos H_0 es decir no hay normalidad.

- Validar Normalidad numéricamente de la PUNT_GLOBAL Variable :

H_0 = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

n = 12374

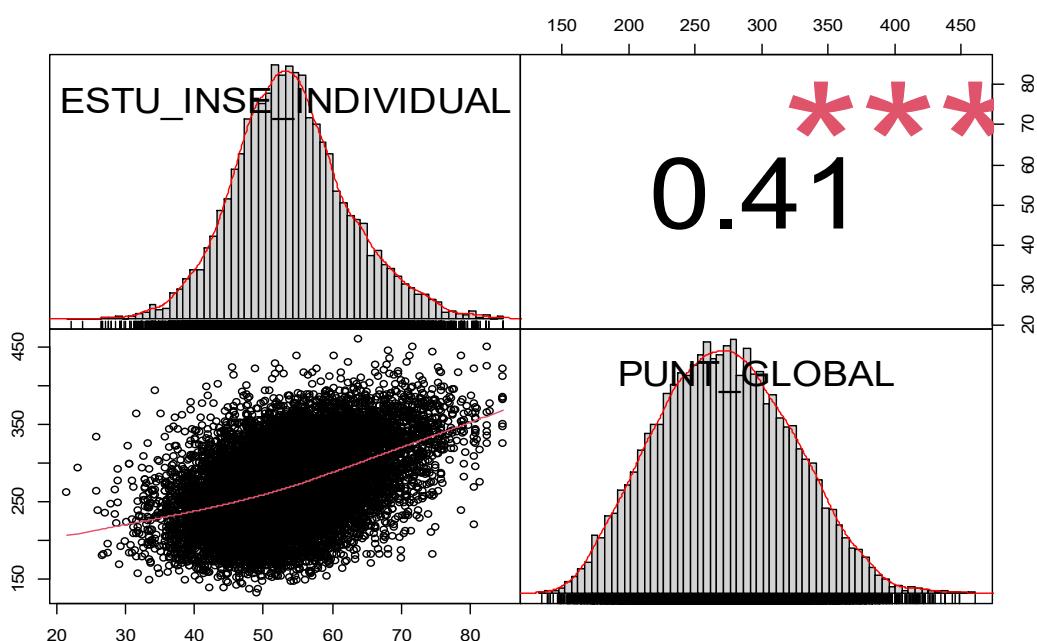
Método a utilizar :
RESULTADOS

KOLMOGOROV - SMIRNOV (KS)
Est. De Prueba (D) P-Valor
0,023264 3,047E-06

Podemos apreciar un P-Valor muy por debajo del Nivel de Significancia ($\alpha = 0,05$) por lo que Rechazamos H_0 es decir no hay normalidad.

F.1 - Análisis grafico de la significancia

Plot de Correlación



- Análisis analítico de la significancia para definir Correlación

Como pudimos constatar en el análisis de normalidad definimos las dos variables como no normales por lo tanto utilizamos el método de Spearman para definir la correlación.

$n =$	12374	$H_0 =$	El coeficiente de Correlación $r = 0$
Método:	Spearman	$H_1 =$	El coeficiente de Correlación $r \neq 0$
		$\alpha =$	0,05
RESULTADOS		Est. De Prueba (S)	P-Valor
		$1,85 \cdot 10^{11}$	$2,2 \cdot 10^{-16}$
			0,4139

Analizamos la correlación a través del método de SPEARMAN, arrojando como resultado un valor P pequeño my pequeño ($x10^{-16}$), es decir rechazamos H_0 dandonos un valor para el coeficiente de Correlación (r) de 0,413, es decir $0.3 < r < 0.5$ estamos en una correlación media.

2

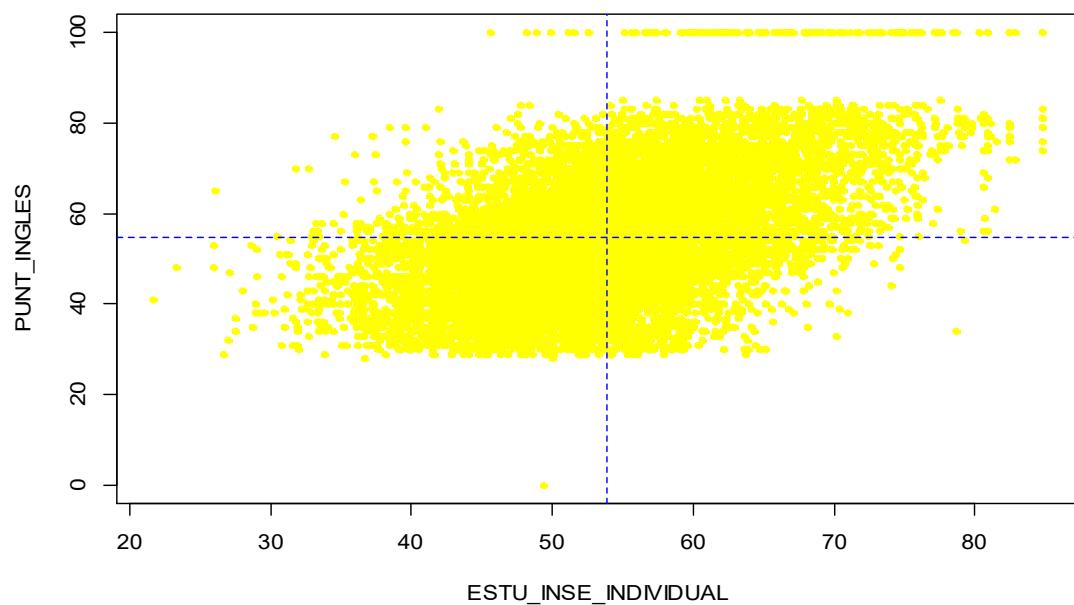
Variable Regresora o Independiente Variable Dependiente

ESTU_INSE_INDIVIDUAL

PUNT_INGLES

A.2

Diagrama Dispersion



B.2

Al parecer hay una relación lineal entre las variables aunque débil.

ESTU_INSE_INDIVIDUAL

PUNT_INGLES

mean

n =

12374

53,85

54,75

C.2

Covarianza

58,39

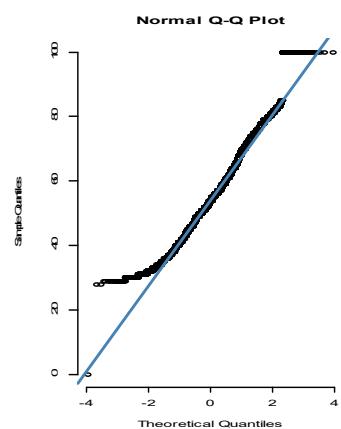
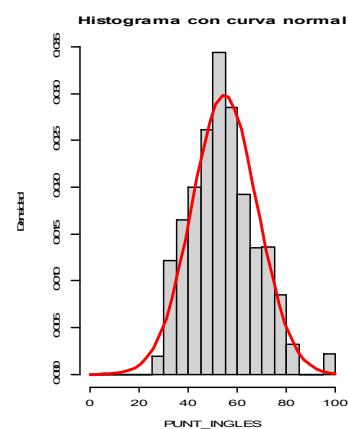
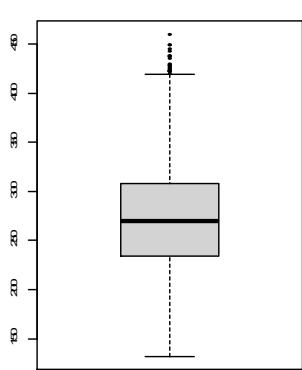
Lo que nos indica la covarianza es que la relación es Directa.

D.2

Validar Normalidad gráficamente de la

PUNT_INGLES

Variable :



Gráficamente podemos apreciar que esta variable tiene un comportamiento aproximadamente normal en los tres gráficos, aunque en el qqplot se ve en la cola derecha e izquierda los valores se están alejando de la media.

- E.2 - Validar Normalidad numéricamente de la Variable : PUNT_INGLES

H_0 = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

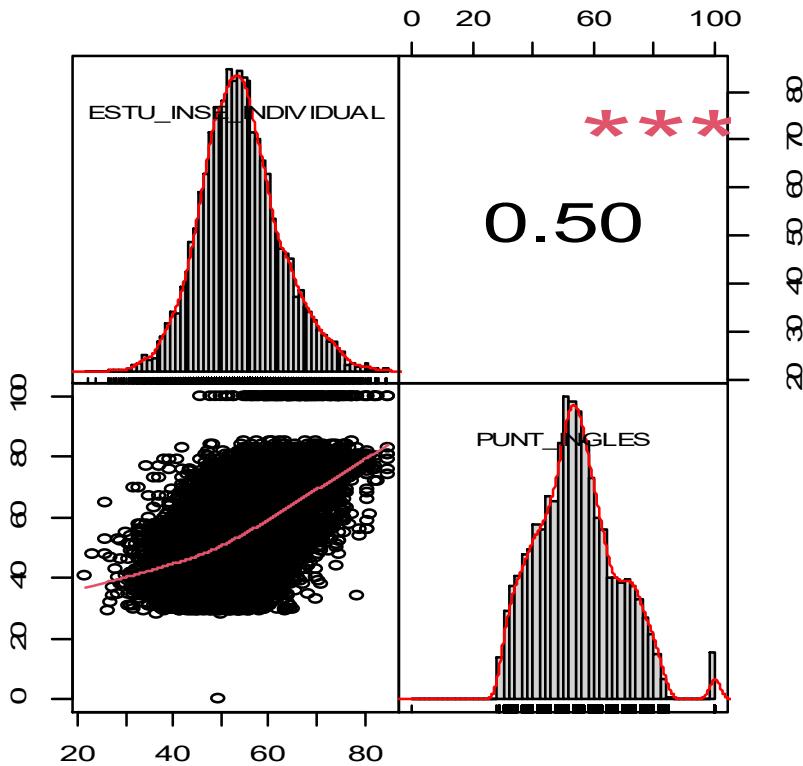
n = 12374

Método a utilizar :	KOLMOGOROV - SMIRNOV (KS)	
RESULTADOS	Est. De Prueba (D)	P-Valor
	0,05214	2,2E-16

Podemos apreciar un P-Valor muy por debajo del Nivel de Significancia ($\alpha = 0,05$) por lo que Rechazamos H_0 es decir no hay normalidad.

- F.2 - Análisis grafico de la significancia

Plot de Correlación



- Análisis analítico de la significancia para definir Correlación

Como pudimos constatar en el análisis de normalidad definimos las dos variables como no normales por lo tanto utilizamos el método de Spearman para definir la correlación.

$n =$	12374	$H_0 =$	El coeficiente de Correlación $r = 0$
Método:	Spearman	$H_1 =$	El coeficiente de Correlación $r \neq 0$
		$\alpha =$	0,05

RESULTADOS	Est. De Prueba (S)	P-Valor	r
	$1,58 * 10^{11}$	2,2E-16	0,497

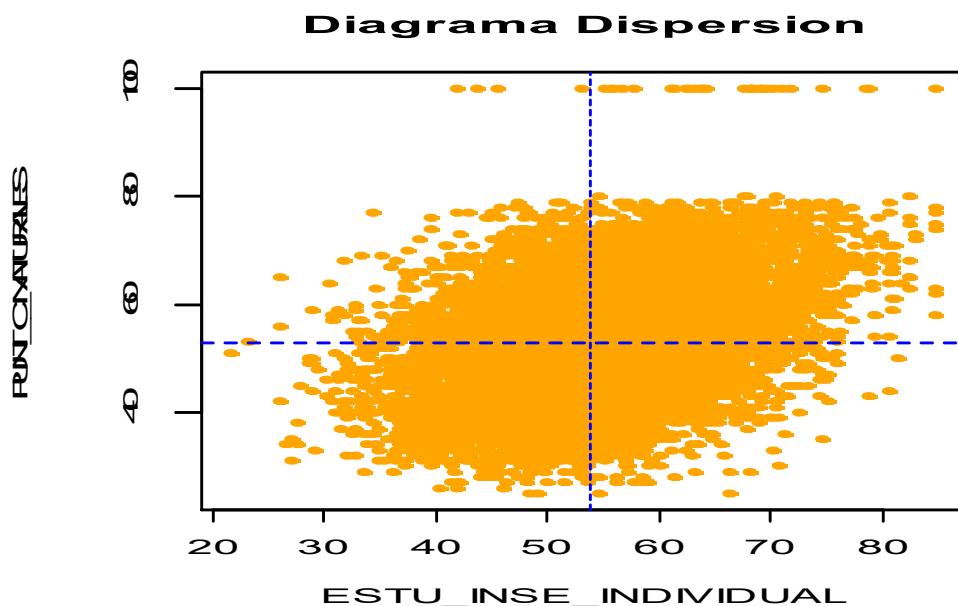
Analizamos la correlación a través del método de SPEARMAN, arrojando como resultado un valor P pequeño y pequeño ($x10^{-16}$), es decir rechazamos H_0 dándonos un valor para el coeficiente de Correlación (r) de 0,4973, es decir $0.3 < r < 0.5$ estamos en una correlación media.

3

Variable Regresora o Independiente Variable Dependiente

ESTU_INSE_INDIVIDUAL PUNT_C_NATURALES

A.3



B.3

Al parecer hay una relación lineal entre las variables aunque débil.

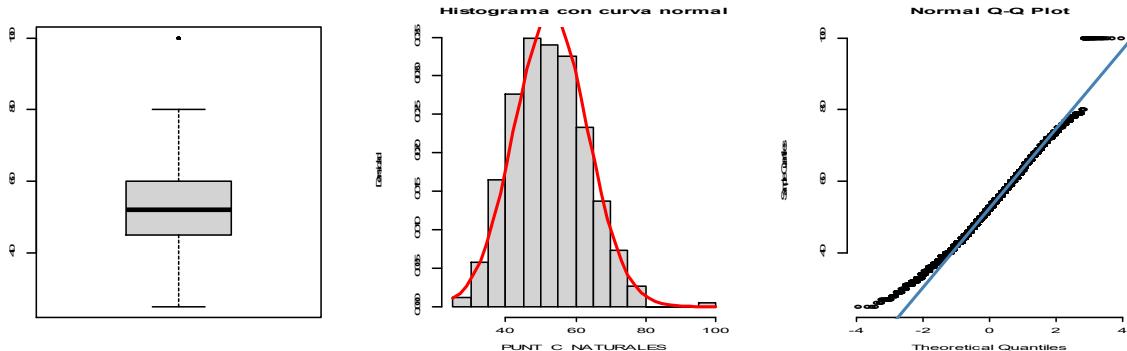
	ESTU_INSE_INDIVIDUAL	PUNT_C_NATURALES
mean	53,85	52,94
$n =$	12374	

C.3

Covarianza 35,54

Lo que nos indica la covarianza es que la relación es Directa.

- D.3 - Validar Normalidad gráficamente de la PUNT_C_NATURALES Variable :



Gráficamente podemos apreciar que esta variable tiene un comportamiento aproximadamente normal en los tres gráficos, aunque en el qqplot se ve en la cola derecha e izquierda los valores se están alejando de la media.

- E.3 - Validar Normalidad numéricamente de la PUNT_C_NATURALES Variable :

H_0 = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

n = 12374

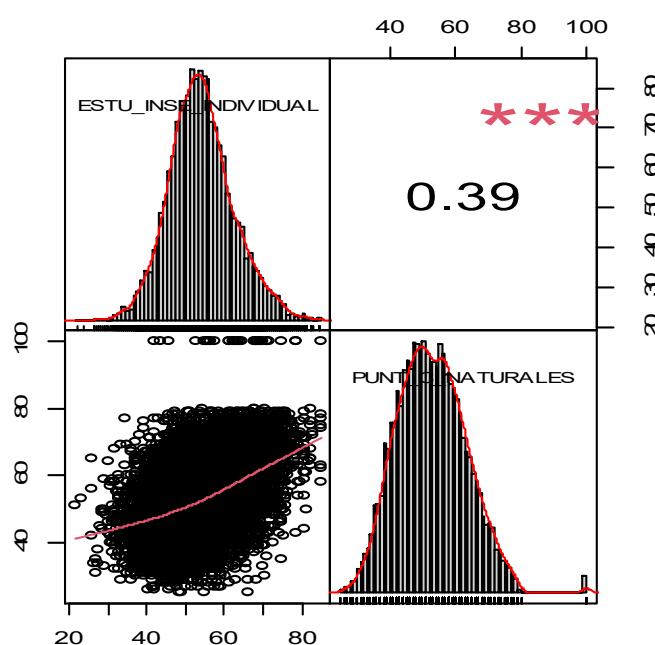
Método a utilizar : KOLMOGOROV - SMIRNOV (KS)

RESULTADOS	Est. De Prueba (D)	P-Valor
	0,0409	2,2E-16

Podemos apreciar un P-Valor muy por debajo del Nivel de Significancia ($\alpha = 0,05$) por lo que Rechazamos H_0 es decir no hay normalidad.

- F.3 - Análisis grafico de la significancia

Plot de Correlación



- Análisis analítico de la significancia para definir Correlación

Como pudimos constatar en el análisis de normalidad definimos las dos variables como no normales por lo tanto utilizamos el método de Spearman para definir la correlación.

$n =$	12374	$H_0 =$	El coeficiente de Correlación $r = 0$
Método:	Spearman	$H_1 =$	El coeficiente de Correlación $r \neq 0$
		$\alpha =$	0,05
RESULTADOS	Est. De Prueba (S)	P-Valor	r
	$1,939 * 10^{11}$	2,2E-16	0,385

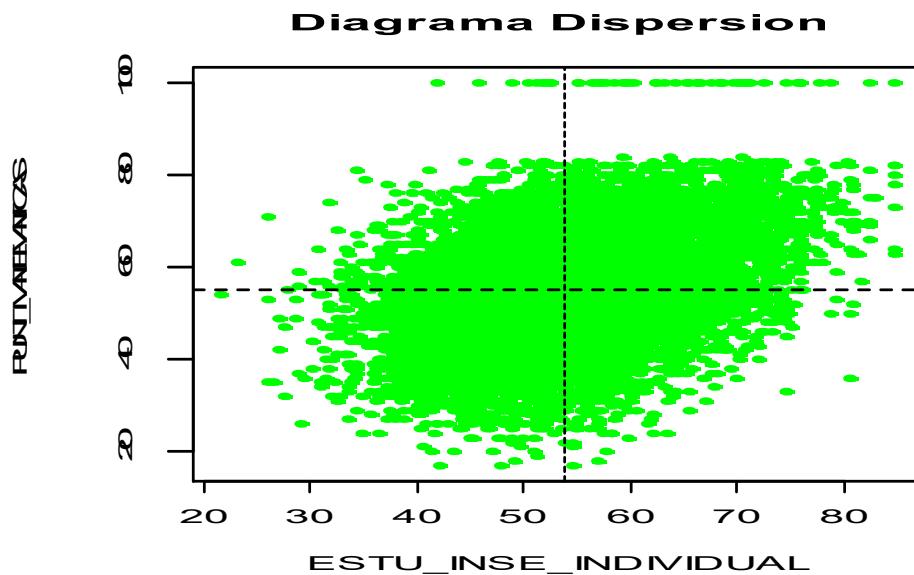
Analizamos la correlación a través del método de SPEARMAN, arrojando como resultado un valor P pequeño y muy pequeño ($x 10^{-16}$), es decir rechazamos H_0 dándonos un valor para el coeficiente de Correlación (r) de 0,385, es decir $0.3 < r < 0.5$ estamos en una correlación media.

4

Variable Regresora o Independiente Variable Dependiente

ESTU_INSE_INDIVIDUAL PUNT_MATEMATICAS

A.4



B.4

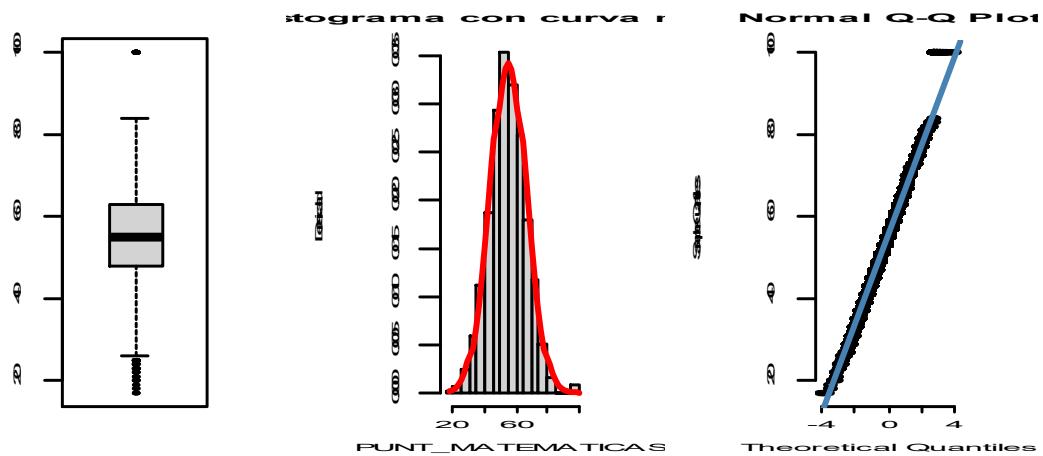
Al parecer hay una relación lineal entre las variables aunque débil.

	ESTU_INSE_INDIVIDUAL	PUNT_MATEMATICAS
mean	53,85	55,21
$n =$	12374	

C.4 Covarianza 37,41

Lo que nos indica la covarianza es que la relación es Directa.

D.4 - Validar Normalidad gráficamente de la PUNT_MATEMATICAS Variable :



Gráficamente podemos apreciar que esta variable tiene un comportamiento aproximadamente normal en los tres gráficos, aunque en el qqplot se ve en la cola derecha valores se están alejando de la media.

E.4 - Validar Normalidad numéricamente de la PUNT_MATEMATICAS Variable :

H_0 = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

n = 12374

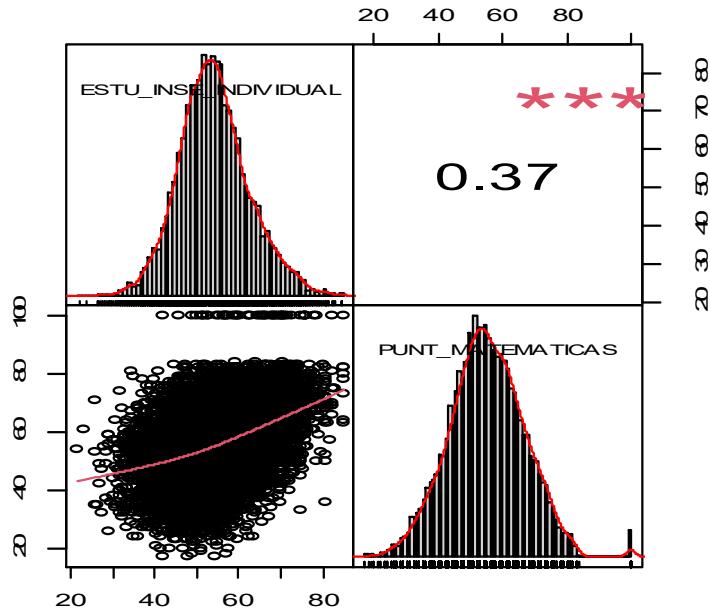
Método a utilizar : KOLMOGOROV - SMIRNOV (KS)

RESULTADOS	Est. De Prueba (D)	P-Valor
	0,0261	9,42E-08

Podemos apreciar un P-Valor muy por debajo del Nivel de Significancia ($\alpha = 0,05$) por lo que Rechazamos H_0 es decir no hay normalidad.

F.4 - Análisis grafico de la significancia

Plot de Correlación



- Análisis analítico de la significancia para definir Correlación

Como pudimos constatar en el análisis de normalidad definimos las dos variables como no normales por lo tanto utilizamos el método de Spearman para definir la correlación.

$n =$	12374	$H_0 =$	El coeficiente de Correlación $r = 0$
Método:	Spearman	$H_1 =$	El coeficiente de Correlación $r \neq 0$
		$\alpha =$	0,05
RESULTADOS	Est. De Prueba (S)	P-Valor	r
	$1,999*10^{11}$	2,2E-16	0,368

Analizamos la correlación a través del método de SPEARMAN, arrojando como resultado un valor P pequeño y pequeño ($x10^{-16}$), es decir rechazamos H_0 dándonos un valor para el coeficiente de Correlación (r) de 0,385, es decir $0.3 < r < 0.5$ estamos en una correlación media.

CONCLUSIONES

- Encontramos como no normales las variables del estudio, por lo cual tuvimos que realizar pruebas no paramétricas (SPEARMAN) para determinar la correlación entre las variables.
- Podemos decir que la correlación entre el INSE y el rendimiento académico es moderada y directa, es decir el nivel socio económico influye o se relaciona directamente con el resultado de las pruebas SABER 11.