

EVALUACIÓN FINAL

Modulo: Estadística y Análisis de Datos

Docente: Andres Jimenez Pacheco

Nombre: Alfonso Camarillo Nùñez

Fecha: 24 de abril de 24

Set de Datos a Usar: créditos_2022

Se anexa la base de datos y el diccionario correspondiente. El objetivo de la evaluación es que sea capaz de crear un reporte estadístico respondiendo preguntas de su interés con la base de datos otorgada, por ende, no deben existir evaluaciones iguales.

Plazo de entrega: Hasta el lunes 06 de mayo de 2024

1. Realizar un reporte descriptivo de las variables de su interés y realizar las interpretaciones correspondientes. Considerar:

- Medidas de Tendencia Central

Estadísticos

	MONTO_CRE DITO	MONTO_VAL OR_VIVIEND A	TASA
N Válido	13507	13507	13507
Perdidos	0	0	0
Media	166270.0604	220434.3812	9.4561
Mediana	148522.0000	205000.0000	8.9000
Moda	119900.00	161000.00	9.50
Suma	2245809705	2977407188	127723.62

Interpretando las medidas de tendencia central tenemos que en promedio los montos crediticios están en \$166,270.06 de estos el 50 % de los montos están en \$148,522.00 o menos y el otro 50% \$148,522.00 o más, y de los montos que se tienen el que más se repite es de \$119,900.00.

De los montos de las viviendas tenemos que en promedio cuestan \$220,434.38 de estos el 50 % de los montos están en \$205,000.00 o menos y el otro 50% \$205,000.00 o más, y de los montos que se tienen el que más se repite es de \$161,000.00.

Sobre la tasa de interés tenemos un promedio de 9.456% el 50 % de las tasas están en 8.9% o menos y el otro 50% 8.9% o más, y de la tasa que más se repite es de 9.5%.

- Medidas de dispersión

Estadísticos

		MONTO_CREDITO	MONTO_VALOR_VIVIENDA	TASA
N	Válido	13507	13507	13507
	Perdidos	0	0	0
Error estándar de la media		835.64500	879.88963	.01783
Desviación estándar		97118.34368	102260.4382	2.07166
Varianza		9431972680	1,046E+10	4.292
Rango		425385.00	416999.52	15.48
Mínimo		4000.00	47200.48	5.00
Máximo		429385.00	464200.00	20.48

	MONTO_CREDITO	MONTO_VALOR_VIVIENDA	TASA
CV	58.4100%	46.3904%	21.9082%

La varianza nos indica la variación, pero al estar en unidades cuadradas es mejor interpretar la desviación estándar que nos indica que la variación entre los montos de crédito es de \$97118.34 y el rango nos indica que la variación entre el monto máximo y mínimo es de \$425,385 la cual oscila entre \$4000 y \$429,385.

Sobre el monto de la vivienda la desviación estándar nos indica que la variación es de \$102,260.43 y el rango nos indica que la variación entre el monto máximo y mínimo es de \$416,999.52 la cual oscila entre \$47,200.48 y \$464,200.00.

En la Tasa tenemos que la desviación estándar nos indica que variación es de 2.07% y el rango nos indica que la variación entre la tasa máxima y mínima es de 15.48% la cual oscila entre 5% y 20.48%.

- Medidas de Posición

Estadísticos

		MONTO_CREDITO	MONTO_VALOR_VIVIENDA	TASA
N	Válido	13507	13507	13507
	Perdidos	0	0	0
Percentiles	25	89600.0000	139000.0000	8.2000
	50	148522.0000	205000.0000	8.9000
	75	235320.0000	295747.0000	9.8000

De esta información tenemos que el primer cuartil de los datos de monto crediticio es de hasta \$89,600 y el tercer cuartil es de hasta \$235,320 por lo que el rango intercuartil es de \$145,720.00 que es a variación en la mitad central de los datos.

Del monto de vivienda el primer cuartil es de hasta \$138,000 y el tercer cuartil es de hasta \$295,747 por lo que el rango intercuartil es de \$156,747.00 que es a variación en la mitad central de los datos.

De la tasa el primer cuartil es de hasta 8.2% y el tercer cuartil es de hasta 9.8% por lo que el rango intercuartil es de 1.6% que es a variación en la mitad central de los datos.

Estadísticos

		MONTO_CREDITO	MONTO_VALOR_VIVIENDA	TASA
N	Válido	13507	13507	13507
	Perdidos	0	0	0
Percentiles	10	47500.0000	92000.0000	7.7000
	20	77600.0000	125240.0000	8.0000
	30	100000.0000	153500.0000	8.4000
	40	119900.0000	172815.6000	8.5900
	50	148522.0000	205000.0000	8.9000
	60	178400.0000	232200.0000	9.2000
	70	217200.0000	274400.0000	9.5000
	80	253871.0000	314500.0000	10.1000
	90	313200.0000	374404.0000	12.4200

Al dividir los datos en 10 partes iguales tenemos los deciles

Estadísticos

		MONTO_CREDITO	MONTO_VALOR_VIVIENDA	TASA
N	Válido	13507	13507	13507
	Perdidos	0	0	0
Percentiles	61	182792.8000	236800.0000	9.2000
	62	186600.0000	240400.0000	9.2400
	63	191400.0000	245000.0000	9.3000
	64	195718.6800	249205.8800	9.3000
	65	199401.5680	252700.0000	9.4000
	66	202665.0284	256600.0000	9.4500
	67	206464.0716	260117.0976	9.5000
	68	209728.8520	265000.0000	9.5000
	69	213222.5840	269400.0000	9.5000

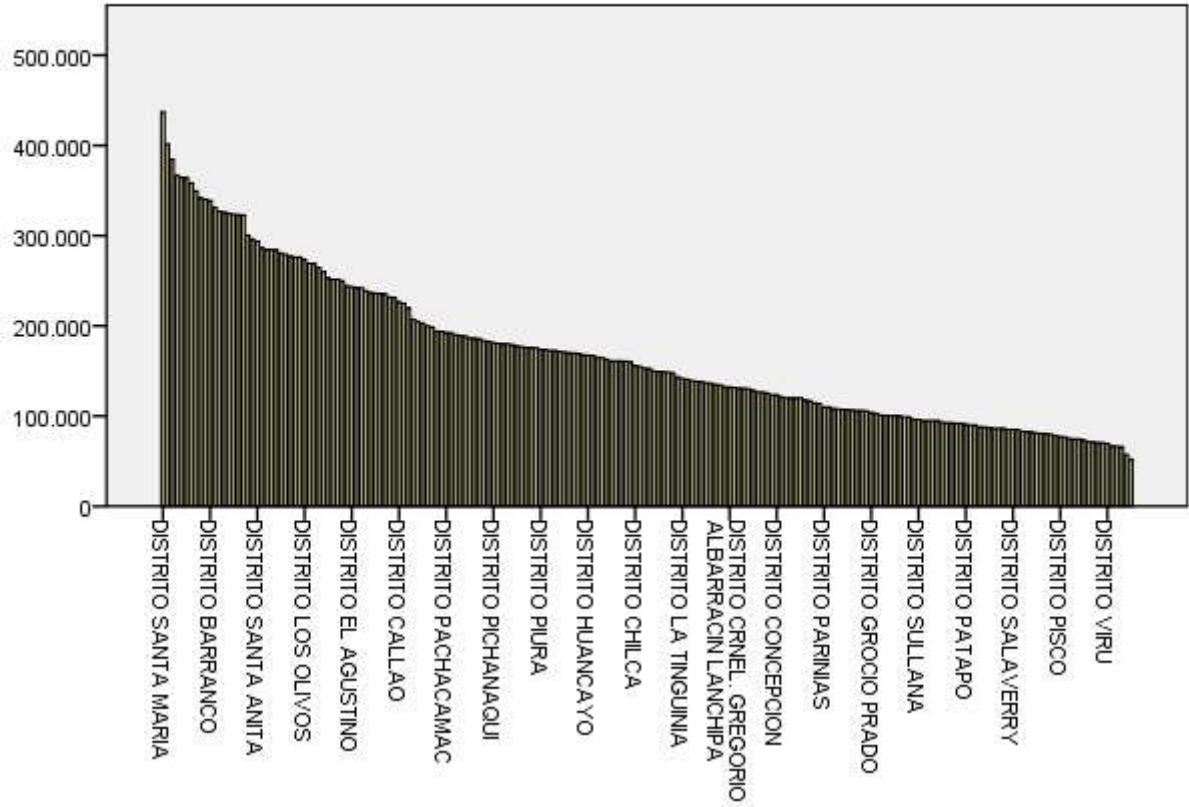
Los percentiles nos indican el valor al dividir los datos en 100 partes iguales por lo que solo se colocan algunos de ejemplo.

2. Formule preguntas de interés con las variables y respóndalas a través de un reporte gráfico. Considere libertad para utilizar las graficas que crea conveniente.

Que distritos tienen los costos promedios de vivienda más altos

		MONTO_VALOR_VIVIENDA
		A
		Media
DISTRITO	SANTA MARIA	437000.00
	LA MOLINA	401500.00
	SAN ISIDRO	384547.02
	SAN BORJA	366566.67
	JESUS MARIA	364426.83
	MAGDALENA DEL MAR	364268.35
	LINCE	358447.77
	SURQUILLO	349573.38
	PUEBLO LIBRE	341846.03
	SANTIAGO DE SURCO	340387.54
	BARRANCO	338999.14
	WANCHAQ	331061.33
	SAN LUIS	327019.24
	SAN MIGUEL	326137.80

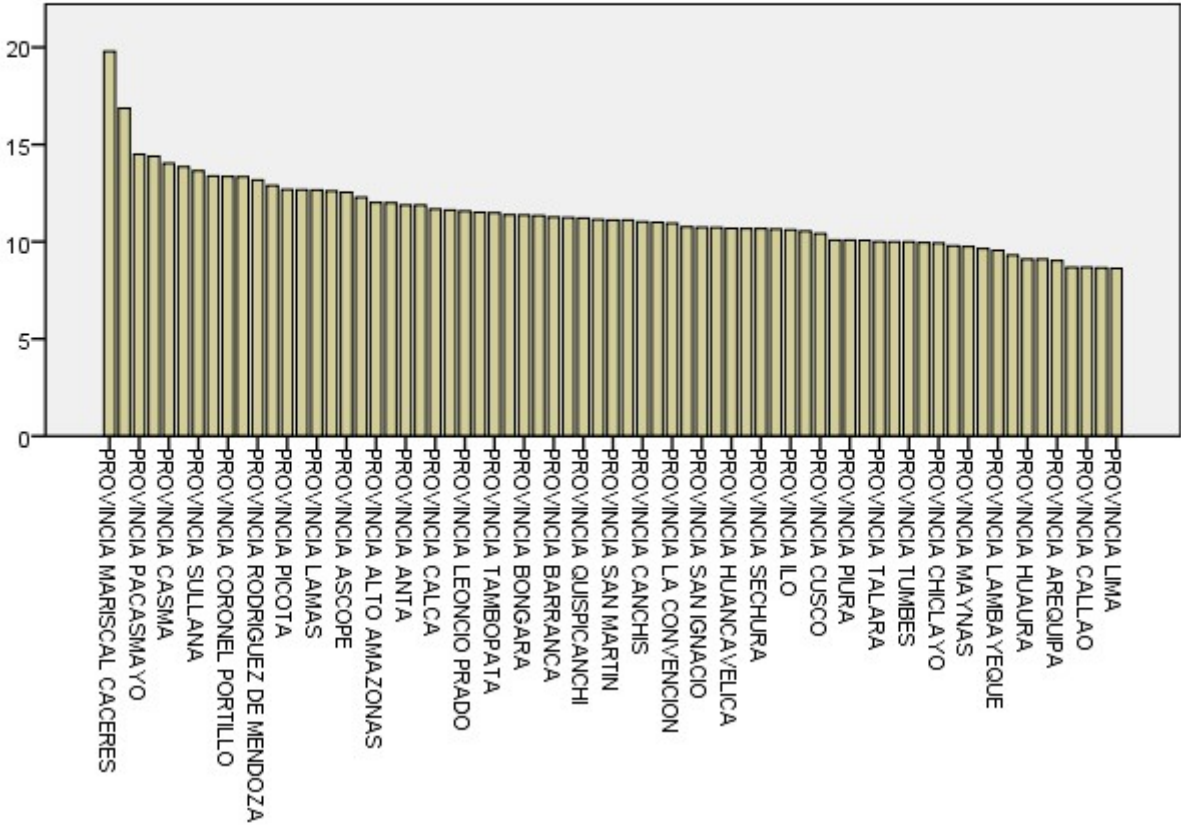
MONTO_VALOR_VIVIENDA Media



Qué provincia tiene en promedio las mayores tasas

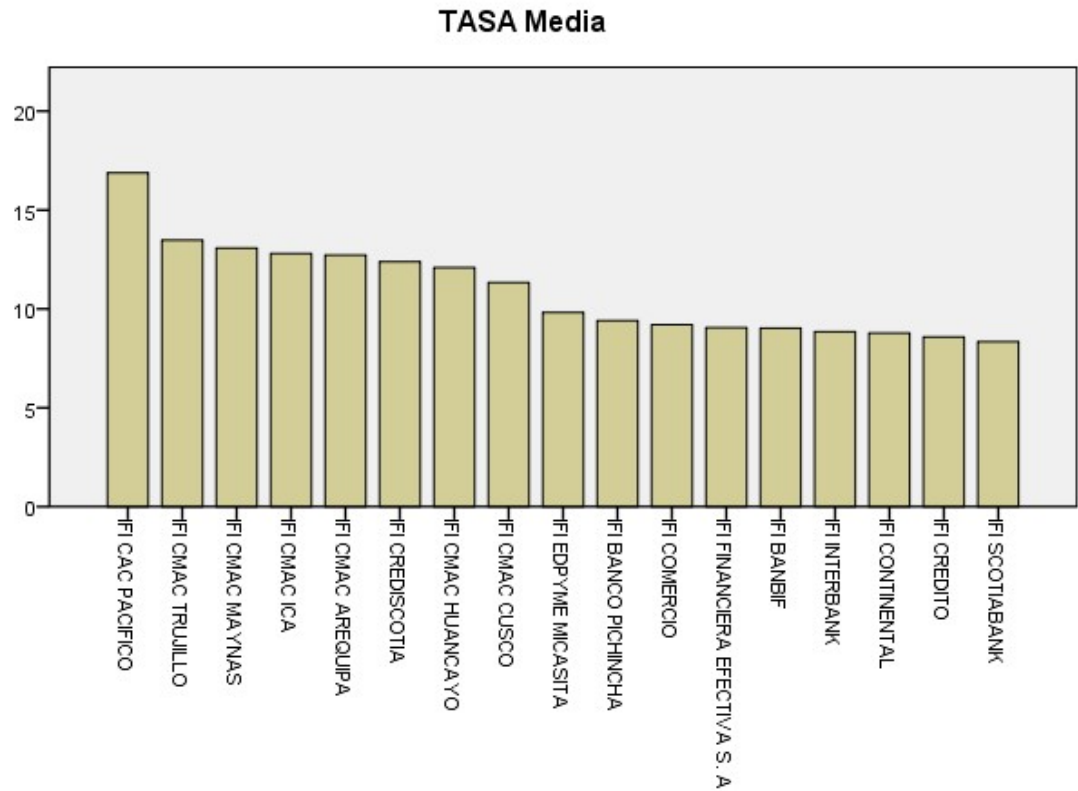
		TASA
		Media
PROVINCIA	MARISCAL CACERES	19.79
	VIRU	16.86
	PACASMAYO	14.50
	MOYOBAMBA	14.39
	CASMA	14.03
	CHINCHA	13.86
	SULLANA	13.65
	PISCO	13.38
	CORONEL PORTILLO	13.35
	CHEPEN	13.33
	RODRIGUEZ DE MENDOZA	13.17
	PAITA	12.89
	PICOTA	12.68
	SATIPO	12.67
	LAMAS	12.66
	RIOJA	12.60
	ASCOPE	12.54

TASA Media

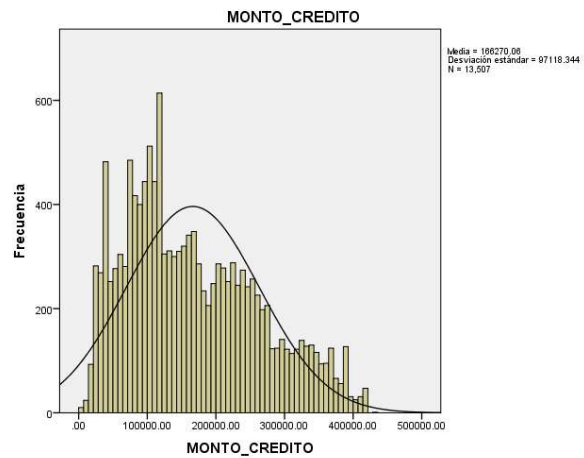


Cuál es la institución financiera que tiene la tasa promedio mas alta

		TASA
		Media
IFI	CAC PACIFICO	16.89
	CMAC TRUJILLO	13.48
	CMAC MAYNAS	13.07
	CMAC ICA	12.80
	CMAC AREQUIPA	12.72
	CREDISCOTIA	12.39
	CMAC HUANCAYO	12.09
	CMAC CUSCO	11.33
	EDPYME MICASITA	9.82
	BANCO PICHINCHA	9.41
	COMERCIO	9.20
	FINANCIERA EFECTIVA S. A	9.05
	BANBIF	9.03
	INTERBANK	8.85
	CONTINENTAL	8.77
	CREDITO	8.58
	SCOTIABANK	8.34



3. ¿Las variables MONTO_CREDITO, MONTO_CUOTA_INICIAL, PLAZOS, ¿TASA siguen una distribución normal?

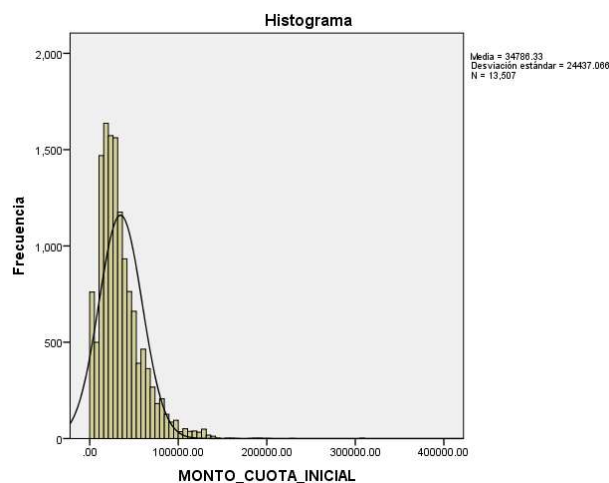


Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
MONTO_CREDITO	.092	13507	.000

a. Corrección de significación de Lilliefors

Al tener un valor menor de 0.05 rechazamos la hipótesis nula por lo que inferimos que no sigue una distribución normal.

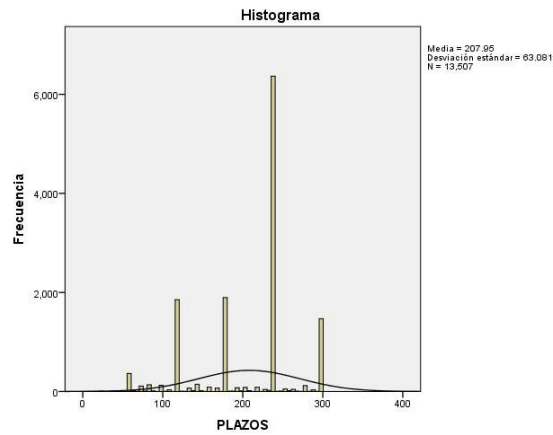


Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
MONTO_CUOTA_INICIAL	.117	13507	.000

a. Corrección de significación de Lilliefors

La grafica podría indicar que si sigue una distribución normal, pero al tener un valor menor de 0.05 rechazamos la hipótesis nula por lo que inferimos que no sigue una distribución normal.

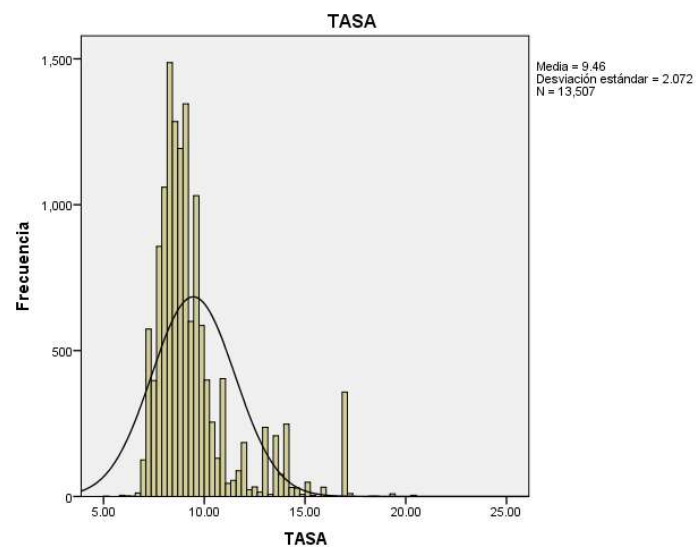


Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
PLAZOS	.292	13507	.000

a. Corrección de significación de Lilliefors

Al tener un valor menor de 0.05 rechazamos la hipótesis nula por lo que inferimos que no sigue una distribución normal.



Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
TASA	.198	13507	.000

a. Corrección de significación de Lilliefors

Al tener un valor menor de 0.05 rechazamos la hipótesis nula por lo que inferimos que no sigue una distribución normal.

4. Si utilizamos la variable MONTO_CUOTA_INICIAL para predecir la variable MONTO_CREDITO ¿Cuál sería la ecuación del modelo y cuáles serían sus medidas de ajuste? ¿Recomendaría este modelo?

El modelo de regresión lineal sería:

$$\text{MONTO_CREDITO} = 95522.108 + 2.034 * (\text{MONTO_CUOTA_INICIAL})$$

Y este modelo explicaría solo un 26.2%, por lo que **no** recomendaría este modelo sería mejor buscar algún otro modelo que explique mejor

Resumen del modelo

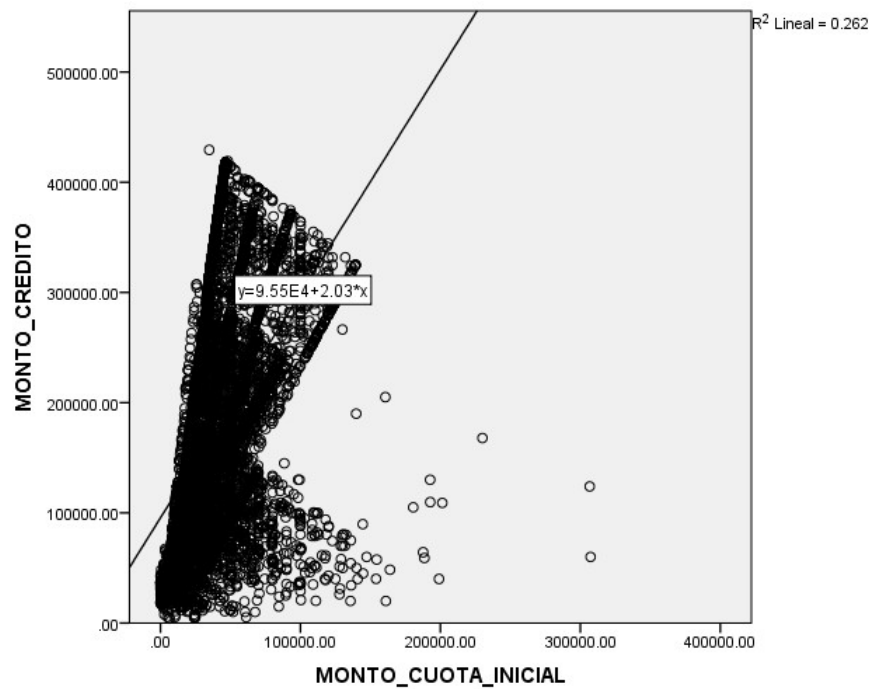
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	.512 ^a	.262	.262	83441.12537

a. Predictores: (Constante), MONTO_CUOTA_INICIAL

Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error estándar	Beta		
1	(Constante)	95522.108	1249.029		76.477	.000
	MONTO_CUOTA_INICIAL	2.034	.029	.512	69.221	.000

a. Variable dependiente: MONTO_CREDITO



Revisando normalidad

Ambas variables son siguen una distribución normal por lo que usaríamos una correlación no paramétrica y tienen una correlación moderada.

Correlaciones

			MONTO_CUOTA_INICIAL	MONTO_CREDITO
Rho de Spearman	MONTO_CUOTA_INICIAL	Coefficiente de correlación	1.000	.631**
		Sig. (bilateral)	.	.000
		N	13507	13507
	MONTO_CREDITO	Coefficiente de correlación	.631**	1.000
		Sig. (bilateral)	.000	.
		N	13507	13507

** . La correlación es significativa en el nivel 0,01 (bilateral).

Prueba de homogeneidad de varianza

		Estadístico de Levene	gl1	gl2	Sig.
MONTO_CUOTA_INICIAL	Se basa en la media	112.091	4	13502	.000
	Se basa en la mediana	69.820	4	13502	.000
	Se basa en la mediana y con gl ajustado	69.820	4	10645.480	.000
	Se basa en la media recortada	90.164	4	13502	.000
MONTO_CREDITO	Se basa en la media	639.203	4	13502	.000
	Se basa en la mediana	605.279	4	13502	.000
	Se basa en la mediana y con gl ajustado	605.279	4	12224.735	.000
	Se basa en la media recortada	637.471	4	13502	.000

Los niveles de significancia son menores al 5% por lo que la varianza no sigue homocedasticidad hay heterocedasticidad.

Si hacemos unas transformaciones la normalidad sigue siendo un problema y que sigue sin tener un valor mayor al 5%.

Pruebas de normalidad

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
Raiz_monto_ini	.049	13506	.000
Log_monto_ini	.086	13506	.000

a. Corrección de significación de Lilliefors

Analizando otros modelos vemos que el que mejor explicaría sería un modelo cubico.

Resumen de modelo y estimaciones de parámetro

Variable dependiente: MONTO_CREDITO

Ecuación	Resumen del modelo					Estimaciones de parámetro			
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Lineal	.262	4791.540	1	13505	.000	95522.108	2.034		
Logarítmico ^a
Cuadrático	.363	3842.203	2	13504	.000	46982.096	4.671	-2.391E-5	
Cúbico	.397	2960.430	3	13503	.000	15360.851	6.979	-6.016E-5	1.308E-10
Exponencial	.240	4254.179	1	13505	.000	82781.776	1.406E-5		

La variable independiente es MONTO_CUOTA_INICIAL.

a. La variable independiente (MONTO_CUOTA_INICIAL) contiene valores no positivos. El valor mínimo es .00. Los modelos Logarítmico y de Potencia no se pueden calcular.

Datos de contacto:

Andres Jimenez

+51 983506650

andresjp.mrk@gmail.com