

## AREMU v2.0

<u>Plataforma:</u>	HP49G
<u>ROM:</u>	1.19-6 B (puede funcionar en anteriores, pero esto no lo he probado)
<u>Programación:</u>	User RPL
<u>Modo de Calculadora:</u>	RPN (pero puede funcionar en Algebraico)
<u>Tipo:</u>	Biblioteca
<u>Número Biblioteca:</u>	1026
<u>Versión:</u>	2.0 (Marzo 2005)
<u>CRC:</u>	#40625d
<u>Tamaño HP49G</u>	10552

### Descripción

AREMu es una programa que calcula las tablas t y ANOVA para regresión lineal múltiple, calcula el coeficiente de correlación, coeficiente de determinación, coeficiente de determinación ajustado, error estándar de estimación, error medio absoluto y estadístico de Durbin-Watson.

### Instalación:

Se debe instalar el programa en un puerto de la calculadora y presionar On+F3 o ejecutar el comando ATTACH (consulte sus manuales para instalar bibliotecas).

### Utilización del programa:

El programa dispone de los siguiente comandos:



**RegLinMult2:** Este comando ejecuta el programa.

**ACERCADE** : Créditos del programa.

RegLinMult2 es el programa que se debe ejecutar. Se puede seleccionar con F1 y aparecerá el siguiente formulario:



El programa toma los datos de la variable **SDAT** si ésta existe, si no existe se pueden ingresar datos nuevos ejecutando el editor de matrices ( $\leftarrow$   $2$  ). Si se ingresan nuevos datos, éstos reemplazarán la variable **SDAT** . Los datos de cada variable corresponden a cada columna ingresada en esta variable.

Si se tienen los datos almacenados en una variable diferente (como los ejemplos suministrados) se puede utilizar la tecla VAR y presionar la tecla de función correspondiente. Y luego presionar OK (F6).

Nota: Los datos de la variable **SDAT** deben estar en modo aproximado, es decir, con un punto al final de cada dato.

El campo ColY se usa para definir la variable dependiente en la matriz **SDAT**, éste valor debe estar entre 1 y el número de columnas de la matriz.

El campo No Dec. Especifica el número de decimales en el reporte generado. El valor predeterminado es 4.

El campo " $\alpha$ " especifica el nivel de confianza en las pruebas estadísticas. Este valor es utilizado para calcular los intervalos de confianza.

Una vez ingresados los datos presionar OK y se verán los resultados, después de los siguientes mensajes:

Calculando.....	Preparando informe...	Calculando ANOVA	Informe ANOVA...
EDIT	EDIT	EDIT	EDIT
CANCEL OK	CANCEL OK	CANCEL OK	CANCEL OK

Calc. Inf. residuos...
EDIT
CANCEL OK

Después que se visualiza el informe aparece el menú de resultados:

Menu de resultados
Ver resultados
Intervalo Predicción
Salir
CANCEL OK

La opción de Ver resultados permite observar los resultados nuevamente. La opción Intervalo Predicción permite calcular un intervalo de predicción al  $100(1-\alpha)\%$ , el valor de  $\alpha$  debe ser introducido en el formulario inicial. La opción Salir permite salir del programa.

### Ejemplo 1

A partir de la siguiente tabla que muestra el comportamiento del salario y la inflación en Colombia entre 1990 y 1998, determinar si existe una relación lineal entre el crecimiento del salario, la inflación y el año.

X1	Y	X2
Año	Crec. Salario	Inflación
1990	26.00	32.36
1991	26.06	26.82
1992	26.04	25.13
1993	25.03	22.60
1994	31.09	22.59
1995	20.05	19.46
1996	19.50	21.63
1997	21.02	17.68
1998	18.05	17.69

- Obtenga una ecuación de regresión que relacione el crecimiento del salario con el año y la inflación.
- Pruebe si existe una relación significativa del crecimiento del salario con las otras dos variables con  $\alpha=0.05$ .

### Solución:

Se ingresan los datos de la siguiente forma:

Si se copiaron los ejemplos, se puede elegir Ejem1 presionando la tecla J y presionando la tecla de función correspondiente. En el campo ColY se digita 2, pues es la columna de la variable y.

Después de digitar los datos y presionar **OK**, se obtendrá el siguiente informe:

#### ---Análisis de Regresión Múltiple---

##### ---Prueba t de regresión---

Intervalo de confianza de los parámetros al 95.2%

Parametro	Estimado	Error Estandar	Estadístico t	Valor p	Lim. Inf.	Lim. Sup.
Constante	3468.2354	2387.7162	1.4525	.1966	-2374.2956	9310.7664
X1.	-1.7229	1.19	-1.4477	.1978	-4.6348	1.189
X2.	-.4006	.6932	-.5779	.5844	-2.0968	1.2956

##### ---Análisis de varianza (ANOVA)---

Fuente	Suma de cuadrados	GL	Medio de cuadrados	F-radio	Valor p
Modelo	74.3245	2.	37.1623	3.2582	.1102
Residual	68.4355	6.	11.4059	0.	0.
Total	142.7601	8.	0.	0.	0.

Coef. de determinación  $R^2= 52.0626\%$   
 Coef. de correlación  $R= .7215$   
 $R^2$  Ajustado= 36.0834%  
 Error est. de la est.= 3.3773  
 Error Medio absoluto= 1.7268  
 Estadístico de Durbin-Watson= 2.6589

a) Según el informe, la ecuación de regresión es

$$Y = 3468.2354 - 1.7229 * X1 - 0.4006 * X2$$

b) Como el valor p (en la tabla ANOVA) es mayor que 0.05, no se rechaza  $H_0: \beta_1 = \beta_2 = 0$ , es decir, no hay relación significativa entre el crecimiento del salario y las otras variables al 95% de nivel de confianza.

### **Ejemplo 2**

Se piensa que la potencia eléctrica consumida al mes por una planta química está relacionada con la temperatura ambiente promedio ( $x_1$ ), el número de días del mes ( $x_2$ ), la pureza promedio del producto ( $x_3$ ) y las toneladas de producto producidas ( $x_4$ ). Los datos correspondientes al año pasado son los siguientes:

y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
240	25	24	91	100
236	31	21	90	95
290	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

Los datos se digitan de la siguiente manera:

Los resultados obtenidos son los siguientes:

---Prueba t de regresión---

Intervalo de confianza de los parámetros al 99.2

Parametro	Estimado	Error Estandar	Estadístico t	Valor p	Lim. Inf.	Lim. Sup.
Constante	-102.7132	207.8586	-.4941	.6363	-830.111	624.6845
X1.	.6054	.3689	1.641	.1448	-.6856	1.8963
X2.	8.9236	5.3005	1.6835	.1361	-9.6254	27.4727
X3.	1.4375	2.3916	.601	.5668	-6.932	9.8069
X4.	.0136	.7338	.0185	.9857	-2.5544	2.5816

---Análisis de varianza (ANOVA)---

Fuente	SUMA de cuadrados	GL	Media de cuadrados	F-ratio	Valor p
Modelo	4957.2447	4.	1239.3112	5.106	.0303
Residual	1699.0053	7.	242.715	0.	0.
Total	6656.25	11.	0.	0.	0.

Coef. de determinacion R<sup>2</sup>= 74.475%

Coef. de correlacion R= .863

R<sup>2</sup> Ajustado= 59.8893%

Error est. de la est.= 15.5793

Error Medio absoluto= 9.7765

Estadístico de Durbin-Watson= 1.7716

Los resultados muestran el ajuste del modelo de regresión lineal múltiple entre Y y 4 variables independientes. La ecuación del modelo ajustado es:

$$Y = -102.7132 + 0.6054 * X_1 + 8.9236 * X_2 + 1.4375 * X_3 + 0.0136 * X_4$$

Para cada una de las siguientes conclusiones suponemos un  $\alpha = 0.05$ .

Los valores de  $f_0=5.106$  y  $p=0.0303$  muestran que se rechaza la hipótesis nula  $H_0: \beta_1 = \beta_2 = \beta_3 = 0$  al nivel de confianza del 95% ( $\alpha=0.05$ ), ya que  $p < \alpha$ . Esto quiere decir que hay evidencia de que hay una relación de manera lineal de la potencia eléctrica consumida mensualmente por la planta eléctrica con las otras variables al 95%.

El coeficiente de determinación  $R^2$  muestra que el modelo ajustado explica el 74.475 % de la variabilidad en Y. El coeficiente de determinación ajustado ( $R^2$  Ajustado en el informe), el cual es más apropiado para comparar modelos con diferente número de variables independientes, es 59.8893 %.

El error o residuo cuadrático medio es igual al cuadrado del Error estándar de la estimación. Esto es :  $(15.5793)^2=242.715$  (en la tabla ANOVA se puede observar este valor). Esta es una estimación de la varianza del error.

El error medio absoluto (9.7765) es el valor promedio de los residuales.

El estadístico de Durbin-Watson prueba los residuales para determinar si hay correlación significativa basado en el orden en el cual ocurren los datos. Como el valor es mayor que 1.4, hay probabilidad de que no hay autocorrelación seria en los residuales.

La prueba t se utiliza para evaluar la contribución al modelo de cada variable. La hipótesis nula es :  $H_0: \beta_j = 0$  contra la hipótesis alternativa:  $H_1: \beta_j \neq 0$ . Como  $p > \alpha$  para cada variable independiente, no se rechaza  $H_0$  y se concluye que cada variable no tiene una contribución importante en el modelo. En la determinación de si el modelo puede ser simplificado, el valor más alto de p en las variables independientes es 0.9857 correspondiente a la variable X4, siendo esta la variable a considerar primero al momento de remover variables del modelo.

Calcular el intervalo de predicción del 95% para el consumo de energía eléctrica cuando  $X_1=75^\circ F$ ,  $X_2=24$  días,  $X_3=90\%$  y  $X_4=98$  Ton

En el menú de resultados se elige la opción Intervalo Predicción y aparecerá un formulario. Después de rellenar este formulario se obtiene:



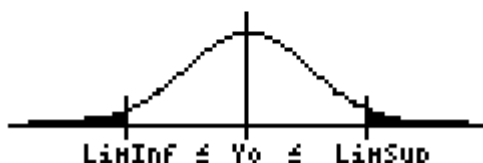
```

Predicción y=a0+a1X1+...+anXn
X1. 75.
X2. 24.
X3. 90.
X4. 98.
Valor de var Ind. X4.
EDIT CANCEL OK
  
```

Luego de presionar OK se obtiene el siguiente informe:

```

=====
Estimación de var. indep.
Yo = 287.5618
=====
Intervalo de predicción al 95.% :
243.7175 ≤ Yo ≤ 331.4062
  
```



**Agradecimientos:**

(en ningún orden específico)

ACO (incluyendo a todos los que participaron en ella) por su trabajo en la HP49G.

Comunidad de AdictosHP por responder todas mis preguntas y CiberEspía, por supuesto.

Elias Valverde por dejarme ver el código fuente de su biblioteca de finanzas.

Eric por mantener su fantástica página / Base de datos: HpCalc.

Dante Camargo S. por su trabajo en LIBKIT v1.1 (utilizada para crear esta biblioteca).

Scott Guth por su biblioteca Stat49 Pro.

Jorge Ayala Niño por su manual de programación (con el que aprendí)

Roger Broncano por su HPUserEdit y manual (algunas veces también lo utilicé)

Dr. Urroz por su contribución al manual de la HP49G+ (con este manual, pude entender mejor la programación gráfica).

Tobias Holgers por su CPictEdit.

A todos los que de momento olvido y que también han escrito manuales.

**Responsabilidad**

No me responsabilizo de los posibles daños que este programa pueda causar a nivel de datos ni a nivel físico en su calculadora.

**Comentarios o sugerencias**

Ing. Carlos Navarro Cera

Ingeniero Industrial Universidad del Atlántico, Barranquilla - Colombia

[carnace@hotmail.com](mailto:carnace@hotmail.com)

**Bibliografía:**

MONTGOMERY, Douglas C. Y RUNGER, George C. Probabilidad y estadística aplicadas a la ingeniería. México : McGrawHill, 1996. 895 p.

**Versiones:**

1.0 Primera versión del programa.

1.01 Corrección de los valores p en la tabla t –había error en los casos en que el estadístico t es negativo- (Thanks, Fabio Frascati).

2.0 Adición de intervalos de confianza para los parámetros de regresión y para predicción. Se pueden estimar nuevos valores.

**Limitaciones:**

Pueden obtenerse resultados diferentes, dependiendo de la relación de dígitos significativos en los datos de las variables. Por ejemplo, en el ejemplo 1 el año se puede cambiar colocando solamente los dos últimos dígitos de este, esto es, para el año 1990 utilizar 90, para el año 1991 utilizar 91, y así sucesivamente; los resultados obtenidos pueden variar en algunas cifras decimales. Esta limitación es debido, en parte, a la precisión de los cálculos utilizados (no tiene nada que ver con el valor ingresado en el formulario de entrada de datos), ya que para elaborar el programa se utilizó métodos matriciales. Para salvar este inconveniente tendría (pienso yo) que pasar el programa a SystemRPL y utilizar los reales extendidos para hacer los cálculos y/o mejorar el algoritmo de cálculo.