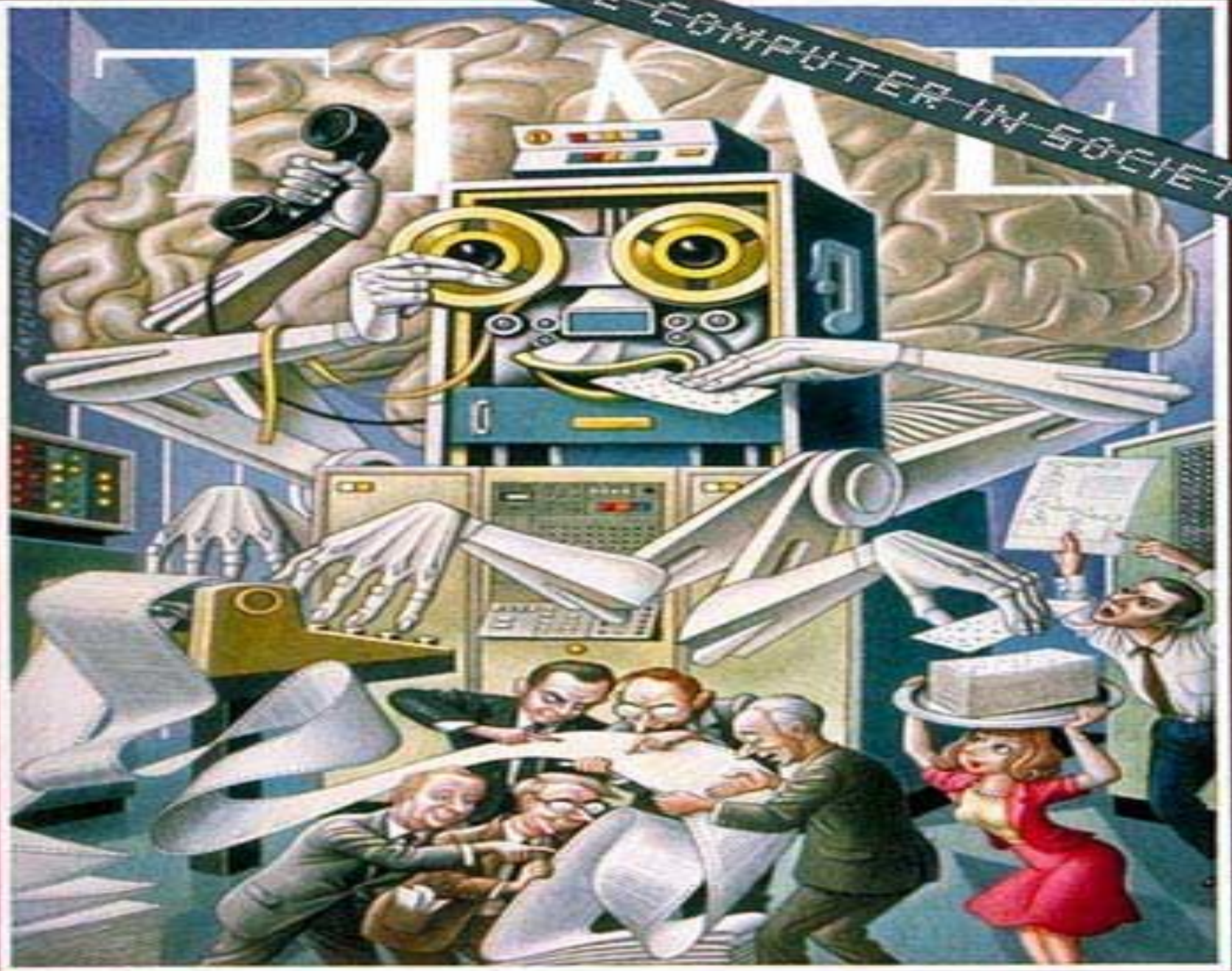


THE COMPUTER IN SOCIETY



Tema 15

Análisis de Regresión Múltiple



Análisis de Regresión Múltiple

- La relación entre las variables no sea lineal
- A que existan muchos casos atípicos o extremos que inciden en la falta de robustez del ajuste de mínimos cuadrados
- A que las variables deban de transformarse para mejorar el ajuste
- A que influyen en la dependiente otras variables no especificadas en el modelo

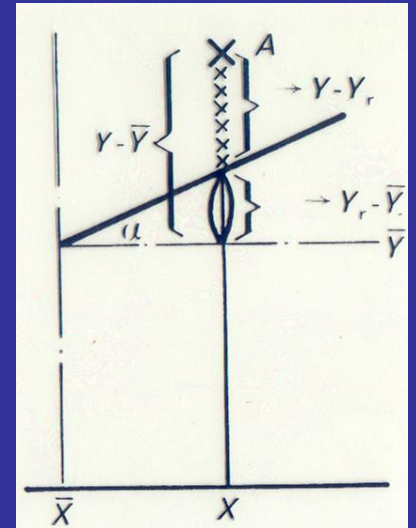


ANÁLISIS DE LOS RESIDUOS

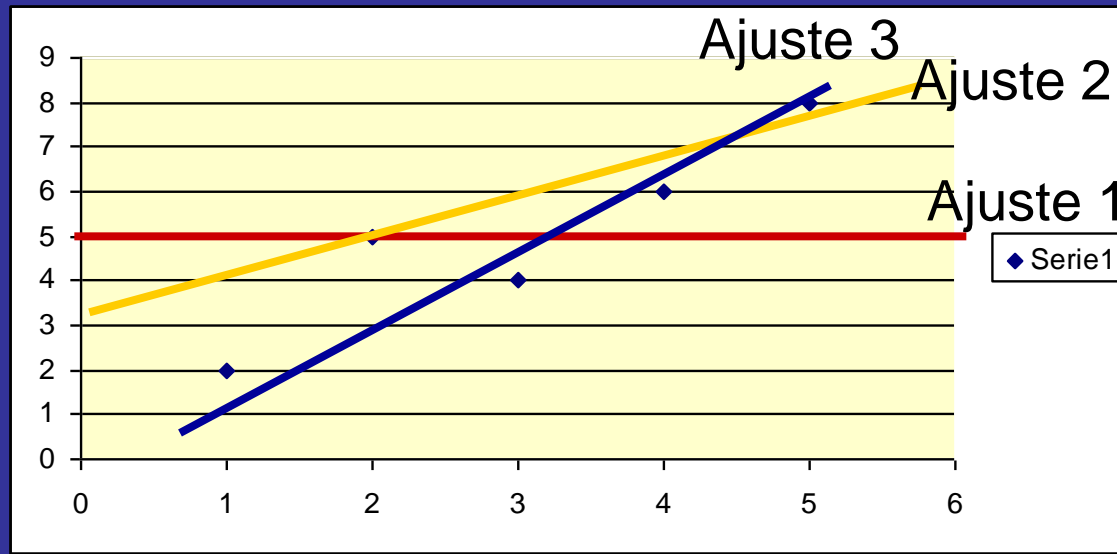
Un **residuo** es la diferencia entre el valor Observado y el estimado a partir del ajuste

Datos = Suave + Aspero

Residuo: Aspero = Datos - suave



ANÁLISIS DE LOS RESIDUOS



Ajuste 1 sin conocer X

$$Y = \bar{Y}$$

Ajuste 2 conociendo X

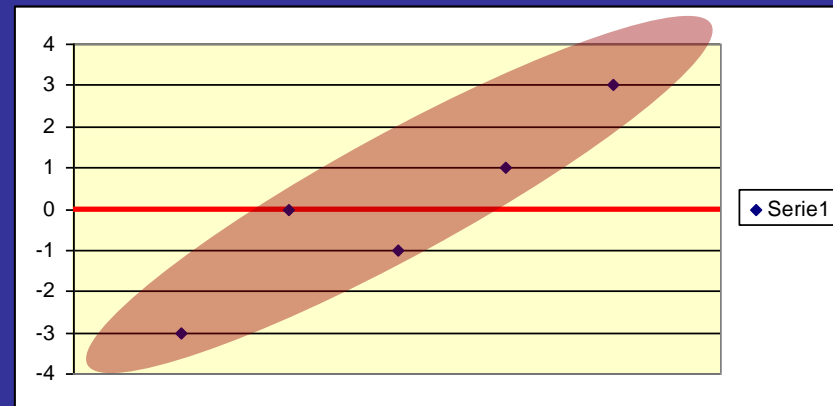
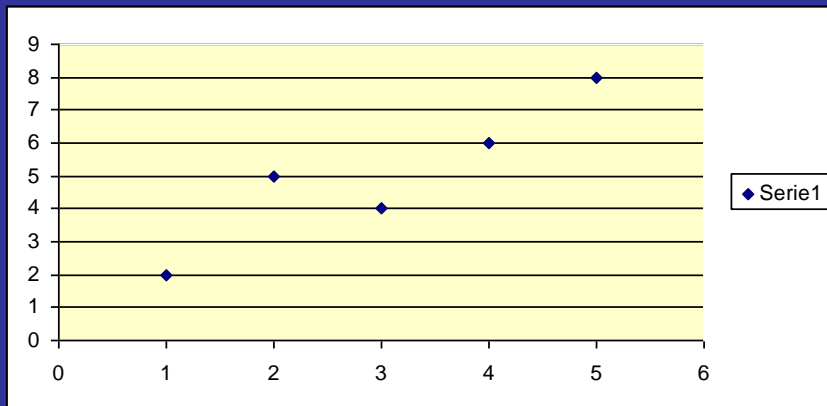
$$Y = 3 + 0,62 X$$

Ajuste 3 conociendo X

$$Y = 1,15 + 1,35 X$$



ANÁLISIS DE LOS RESIDUOS



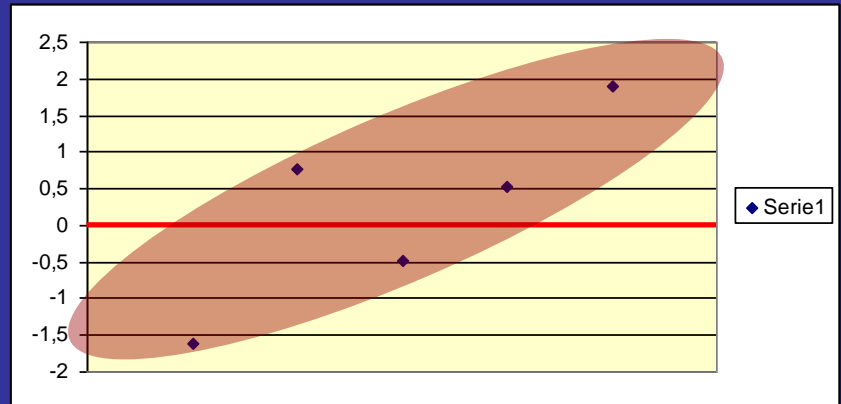
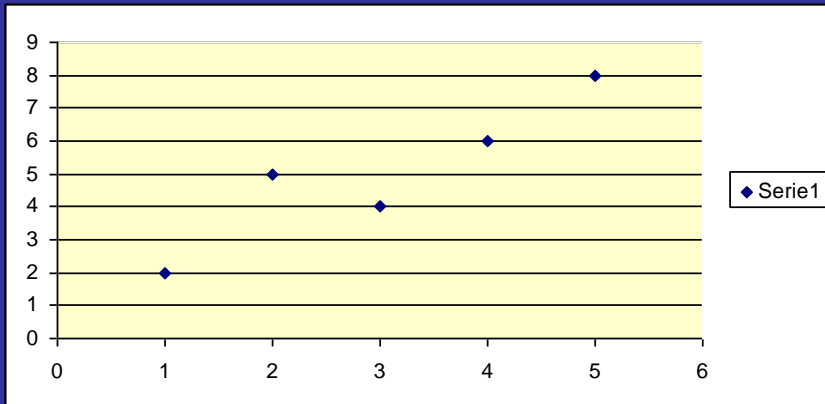
ANÁLISIS DE RESIDUALES

$$(1) Y = \bar{Y}$$

X	Y	Yr	Resid
1	2	5	-3
2	5	5	0
3	4	5	-1
4	6	5	1
5	8	5	3



ANÁLISIS DE LOS RESIDUOS



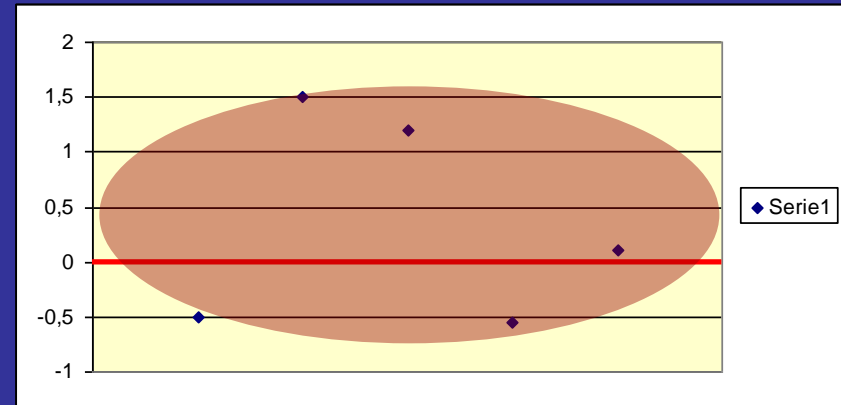
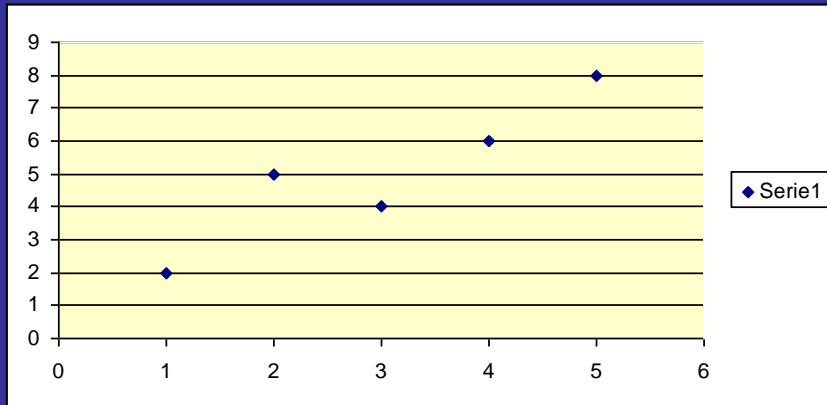
ANÁLISIS DE RESIDUALES

$$(2) Y = 3 + 0,62 X$$

X	Y	Yr	Resid
1	2	3,62	-1,62
2	5	4,24	0,76
3	4	4,86	-0,48
4	6	5,48	0,52
5	8	6,10	1,90



ANÁLISIS DE LOS RESIDUOS



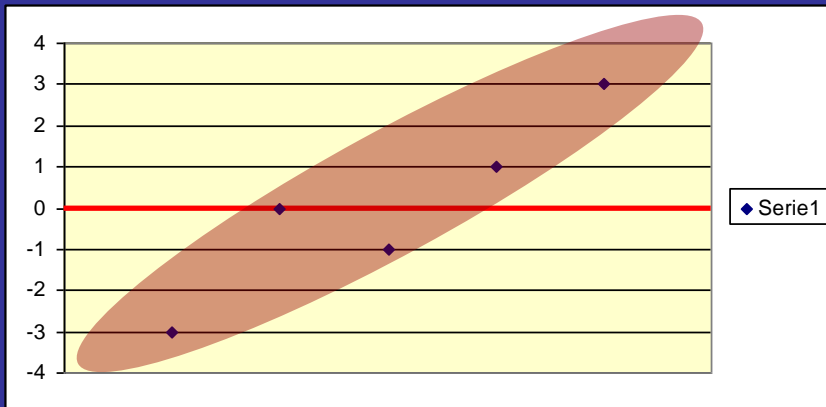
ANÁLISIS DE RESIDUALES

$$(3) Y = 1,15 + 1,35 X$$

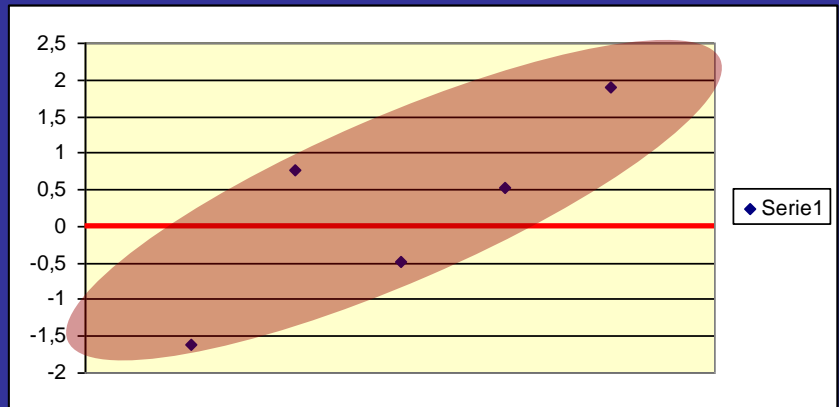
X	Y	Yr	Resid
1	2	2,50	-0,50
2	5	3,85	1,50
3	4	5,20	1,20
4	6	6,55	-0,55
5	8	7,90	0,10



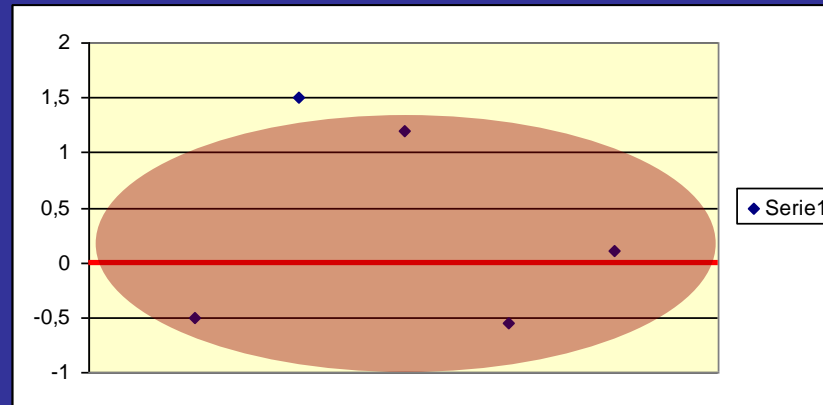
ANÁLISIS DE LOS RESIDUOS



Ajuste 1 sin conocer X $Y = Y$



Ajuste 2 conociendo X $Y = 3 + 0,62 X$



Ajuste 3 conociendo X $Y = 1,15 + 1,35 X$



ANÁLISIS DE LOS RESIDUOS

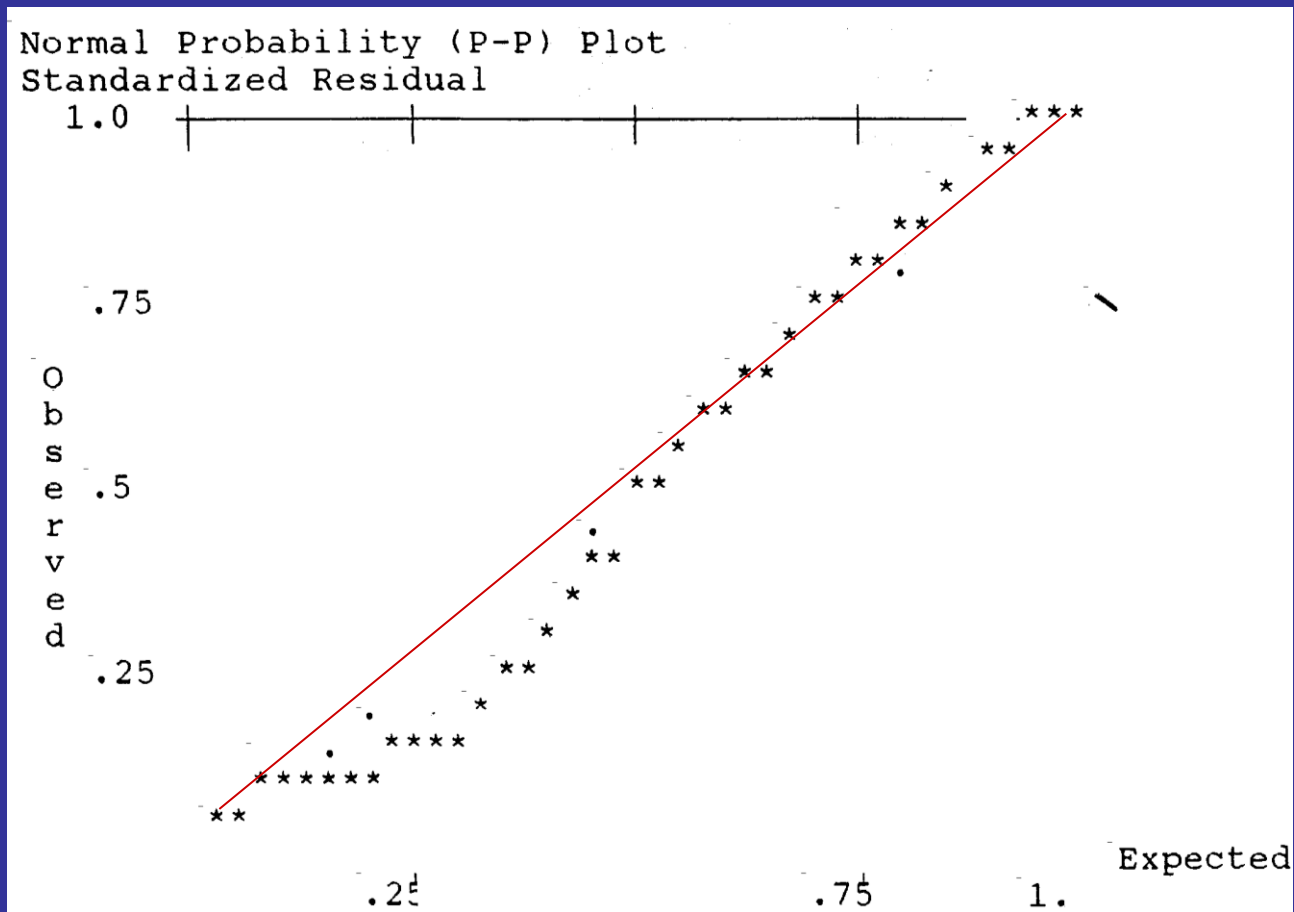
Histogram - Standardized Residual

(* = 1 Cases, = Norma. Curve)

N	Exp N	N	
0	.04	Out	
0	.07	3.00	
0	.18	2.67	
0	.41	2.33	
0	.84	2.00	.
2	1.54	1.67	*:
2	2.52	1.33	**.
4	3.71	1.00	***:
5	4.89	.67	****:
8	5.76	.33	*****: **
10	6.09	.00	*****: *****
6	5.76	-.33	*****:
4	4.89	-.67	****.
1	3.71	-1.00	* .
1	2.52	-1.33	* .
1	1.54	-1.67	*.
0	.84	-2.00	.
0	.41	-2.33	.
1	.18	-2.67	*
0	.07	-3.00	.
1	.04	Out	*



ANÁLISIS DE LOS RESIDUOS



Análisis de Regresión Múltiple

- La relación entre las variables no sea lineal
- A que existan muchos casos atípicos o extremos que inciden en la falta de robustez del ajuste de mínimos cuadrados
- A que las variables deban de transformarse para mejorar el ajuste
- A que influyen en la dependiente otras variables no especificadas en el modelo



Análisis de Regresión Múltiple

- La **transformación** o **reexpresión** de las variables lo que pretende es la consecución de la normalidad en la distribución con el objeto de que se cumplan los supuestos paramétricos y se puedan aplicar sobre las variables este tipo de análisis. Como decimos la transformación persigue la consecución de una distribución aproximada a la normal o al menos simétrica, dado que la mayoría de las técnicas de análisis confirmatorio dan por supuesto esta normalidad o simetría



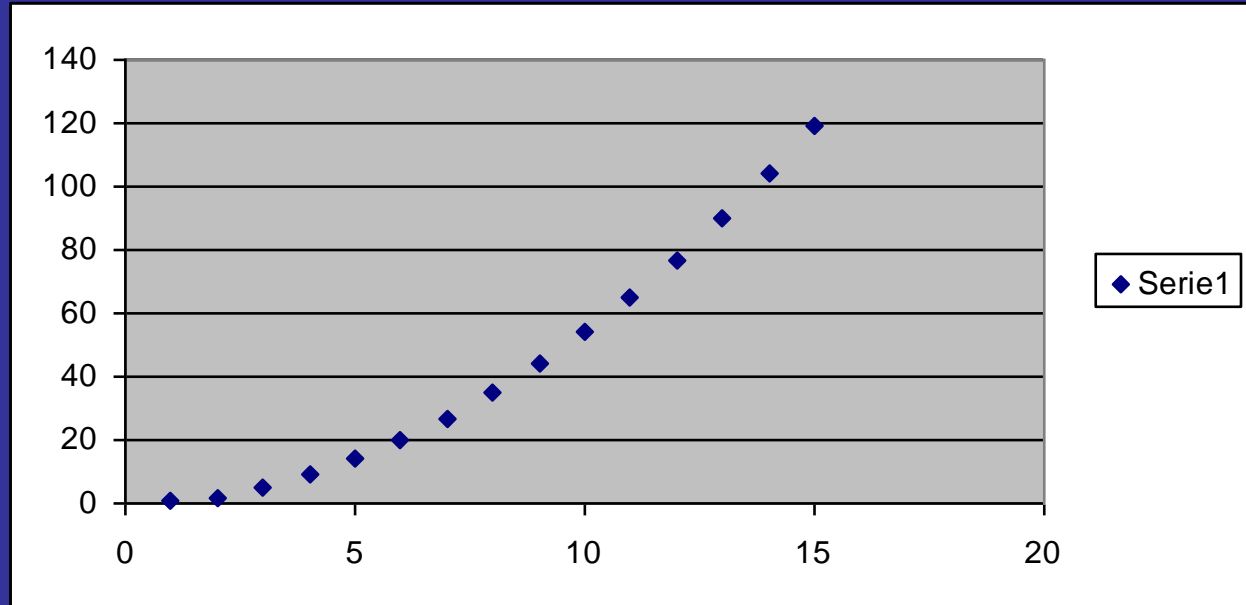
Análisis de Regresión Múltiple

- Para que dos variables estén relacionadas han de tener la misma forma. El hecho de que tengan la misma forma no asegura que estén relacionadas pero no pueden estarlo si sus distribuciones son muy diferentes
- Es decir hay una relación directa entre normalidad y linealidad
- Si una variable no se distribuye normalmente, existen una serie de transformaciones que la acercan a la normalidad y con ello se favorece la asociación lineal



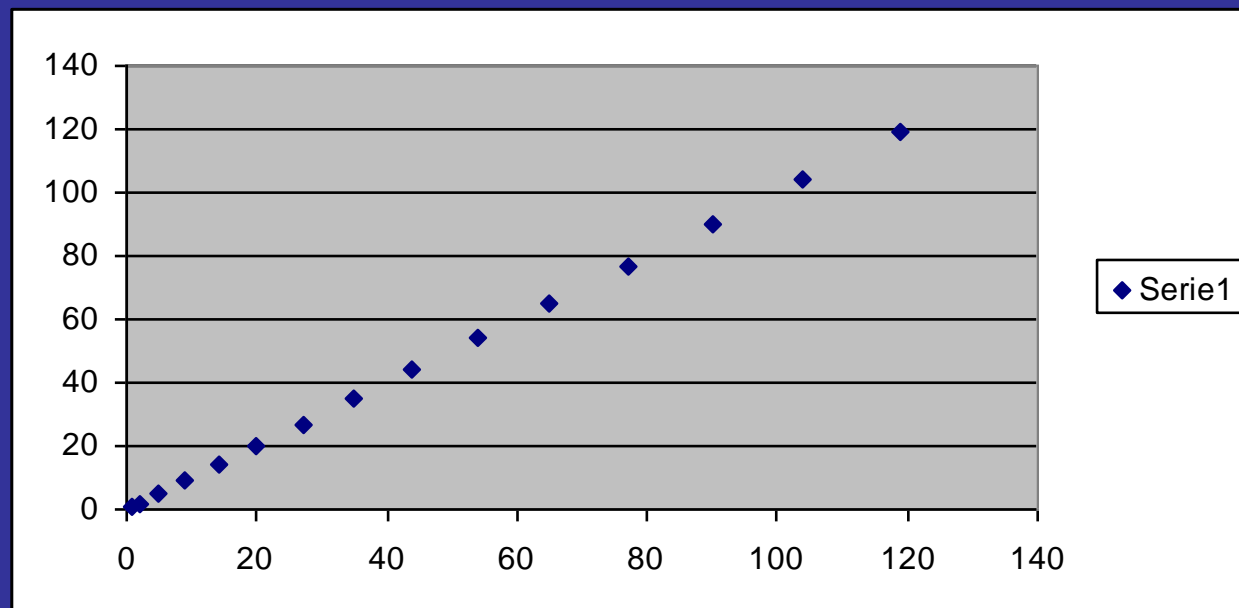
Análisis de Regresión Múltiple

1	1
2	2
3	5
4	9
5	14
6	20
7	27
8	35
9	44
10	54
11	65
12	77
13	90
14	104
15	119



Análisis de Regresión Múltiple

1 1
2 2
5 5
9 9
14 14
20 20
27 27
35 35
44 44
54 54
65 65
77 77
90 90
104 104
119 119



Análisis de Regresión Múltiple

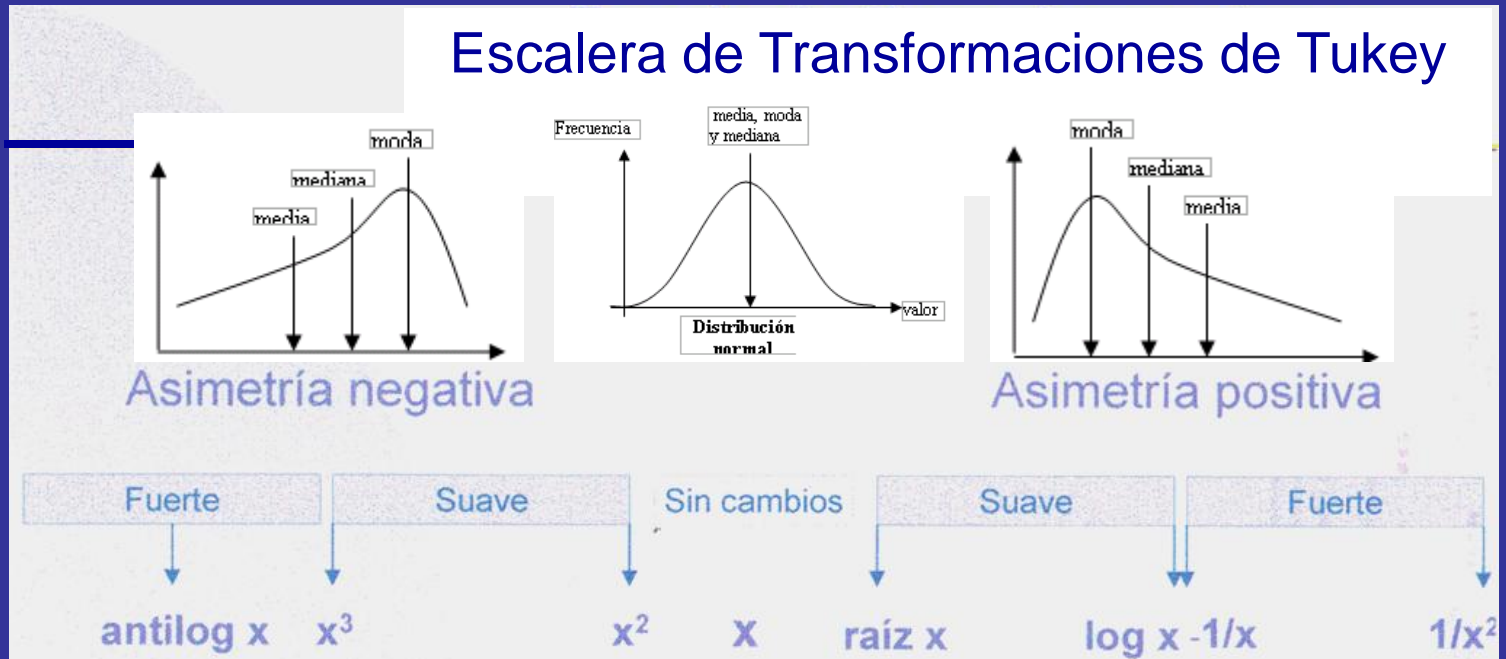
TIPOS DE TRANSFORMACIONES

- **Lineal** (Suma/Resta/Multiplicación/División)
Únicamente cambia los valores brutos de la variable no cambia la distancia, el orden ni la forma de la distribución
- **No Lineal monotonica** (raíces, logaritmos, potencias)
Cambia el valor, la distancia entre los valores pero no el orden ($y = a + b x^2$)
- **No lineal no monotonica** ($y = a + b_1 X + b_2 X^2 + b_3 X^3 + \dots + b_n X^n$)
Cambia el valor, la distancia y el orden

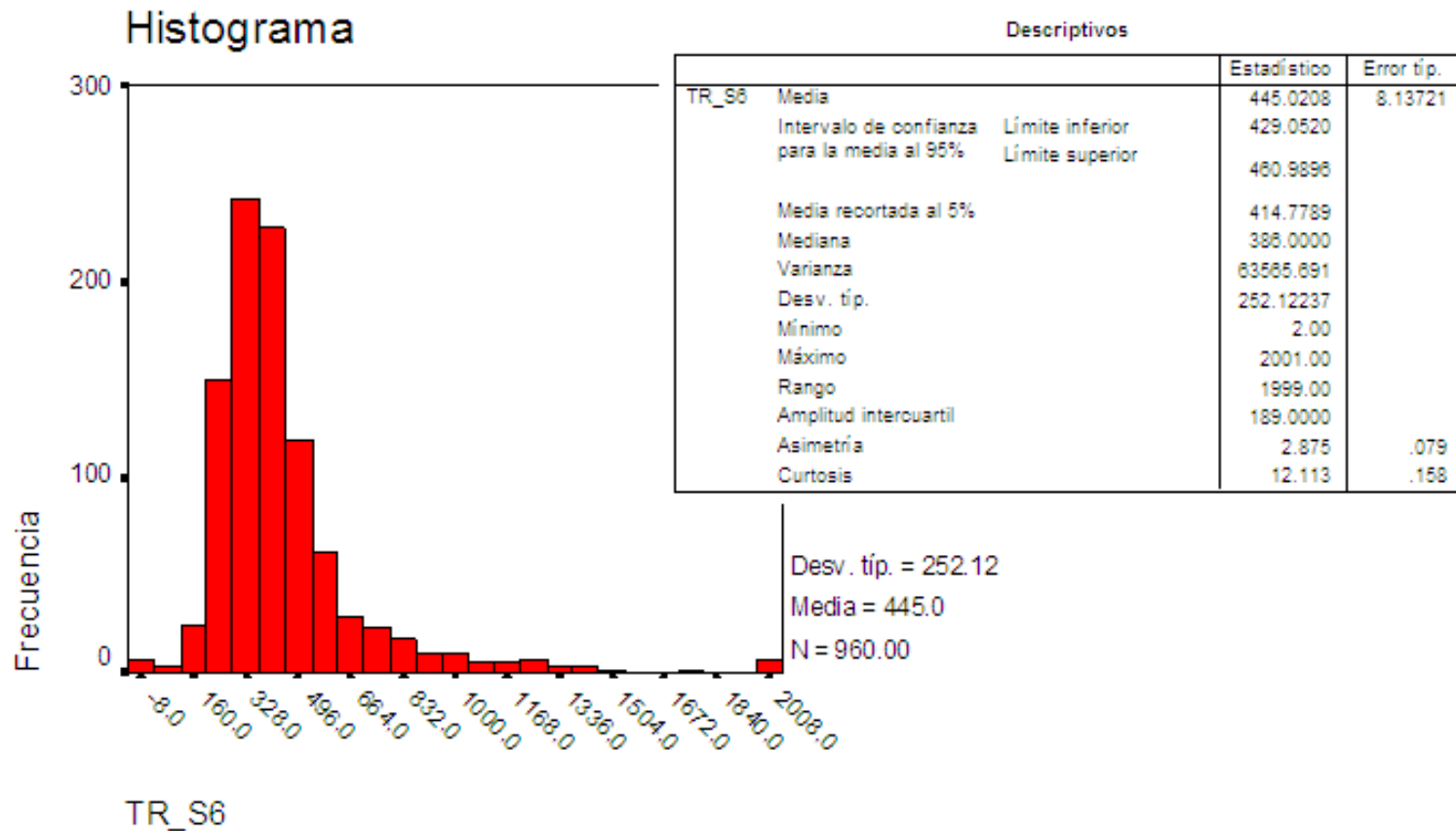


Análisis de Regresión Múltiple

Escalera de Transformaciones de Tukey

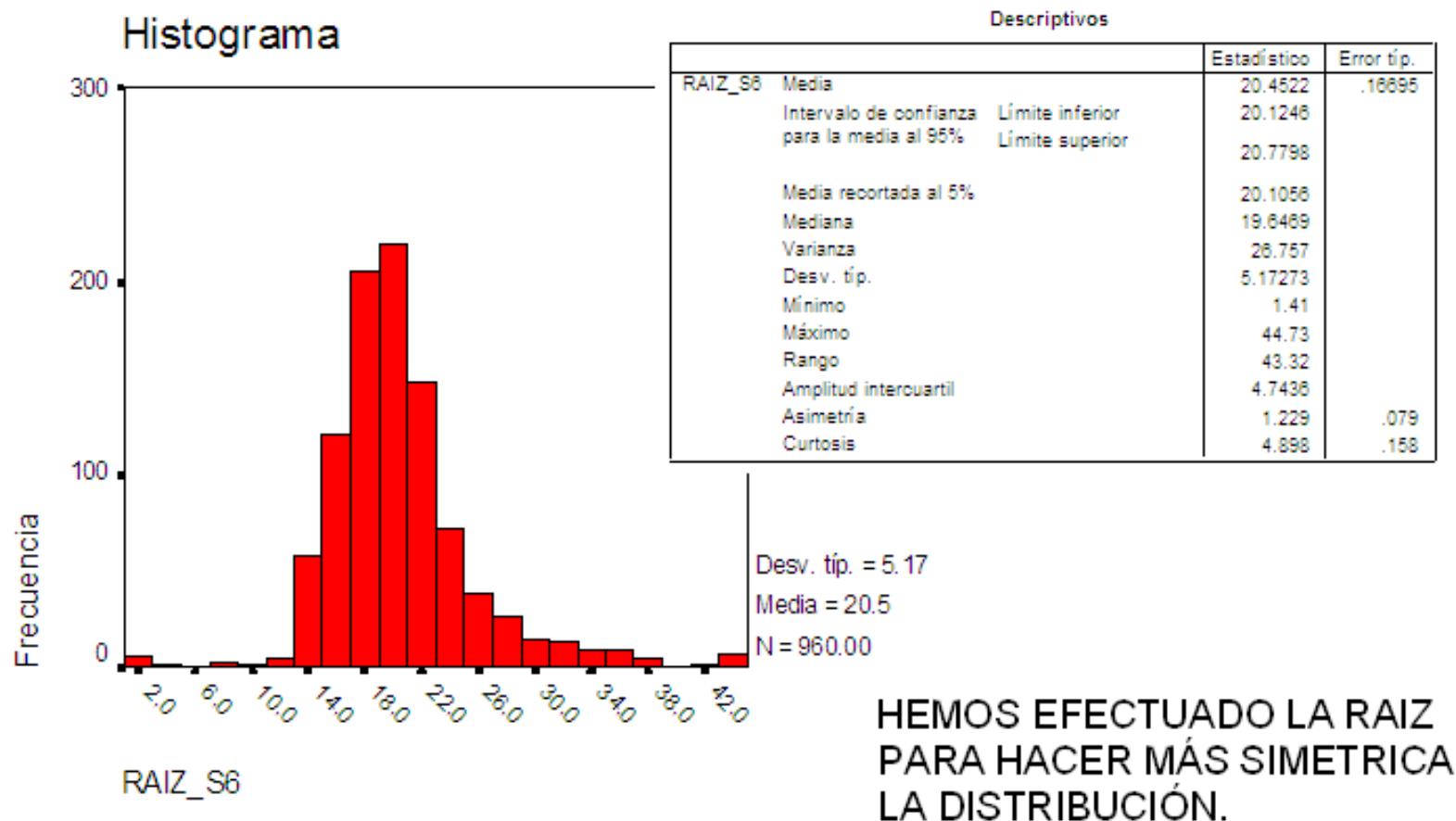


Ejemplo. Datos de TR de un participante



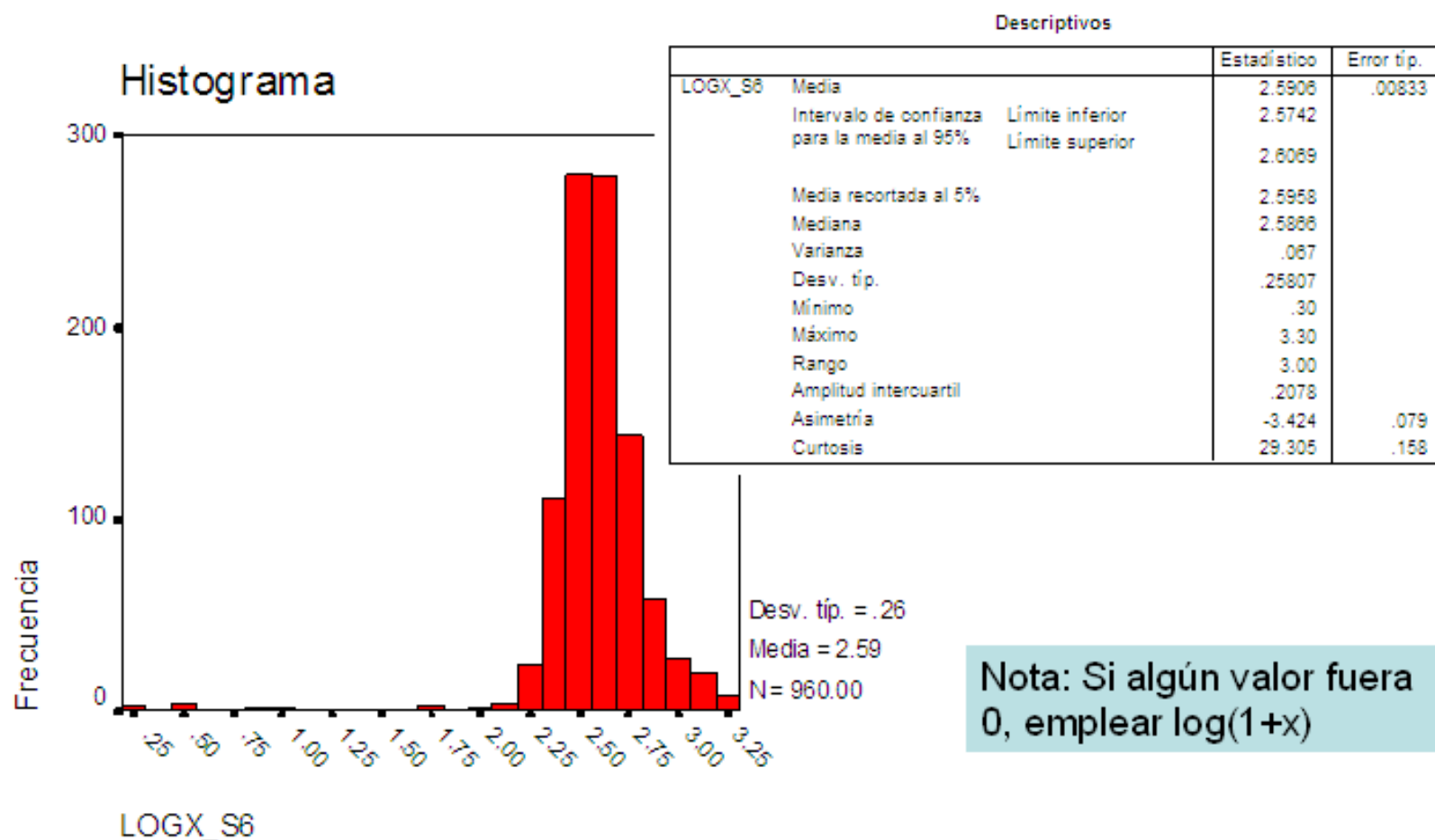
Observad no sólo que hay algunas puntuaciones atípicas a ambos lados, sino que hay una clara asimetría positiva.

Ejemplo. Datos (transformados; raíz cuadrada) de TR de un participante (cont.)



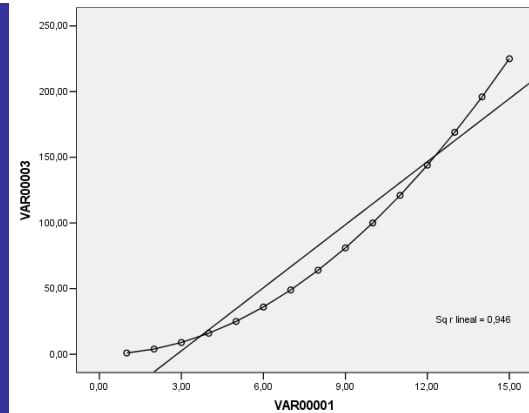
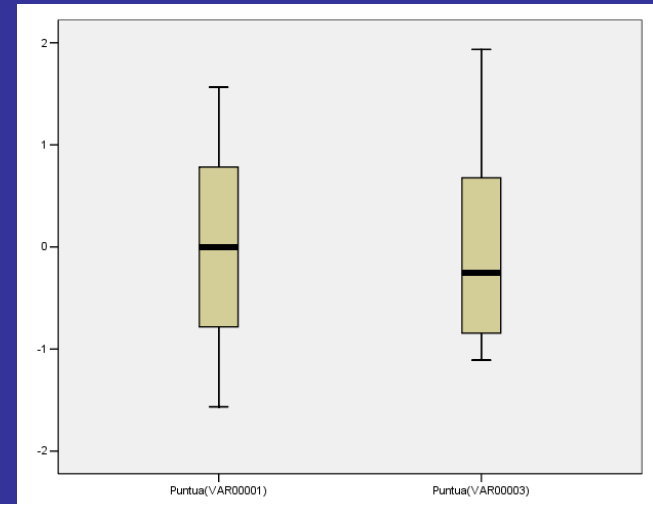
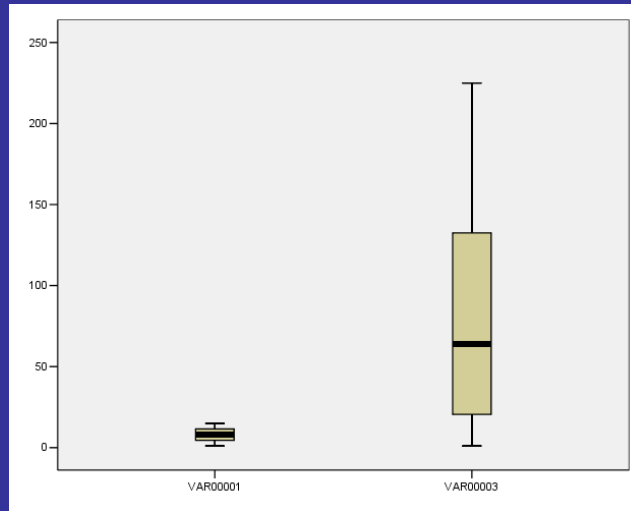
Observad no sólo que aún queda algo de asimetría positiva. Con el logaritmo, podremos reducir más la asimetría positiva, es lo que haremos ahora

Ejemplo. Datos (transformados; logaritmo) de TR de un participante (cont.)



Observad no sólo que la asimetría positiva ha desaparecido (si acaso hay cierta asimetría negativa causada por unas pocas puntuaciones atípicas).

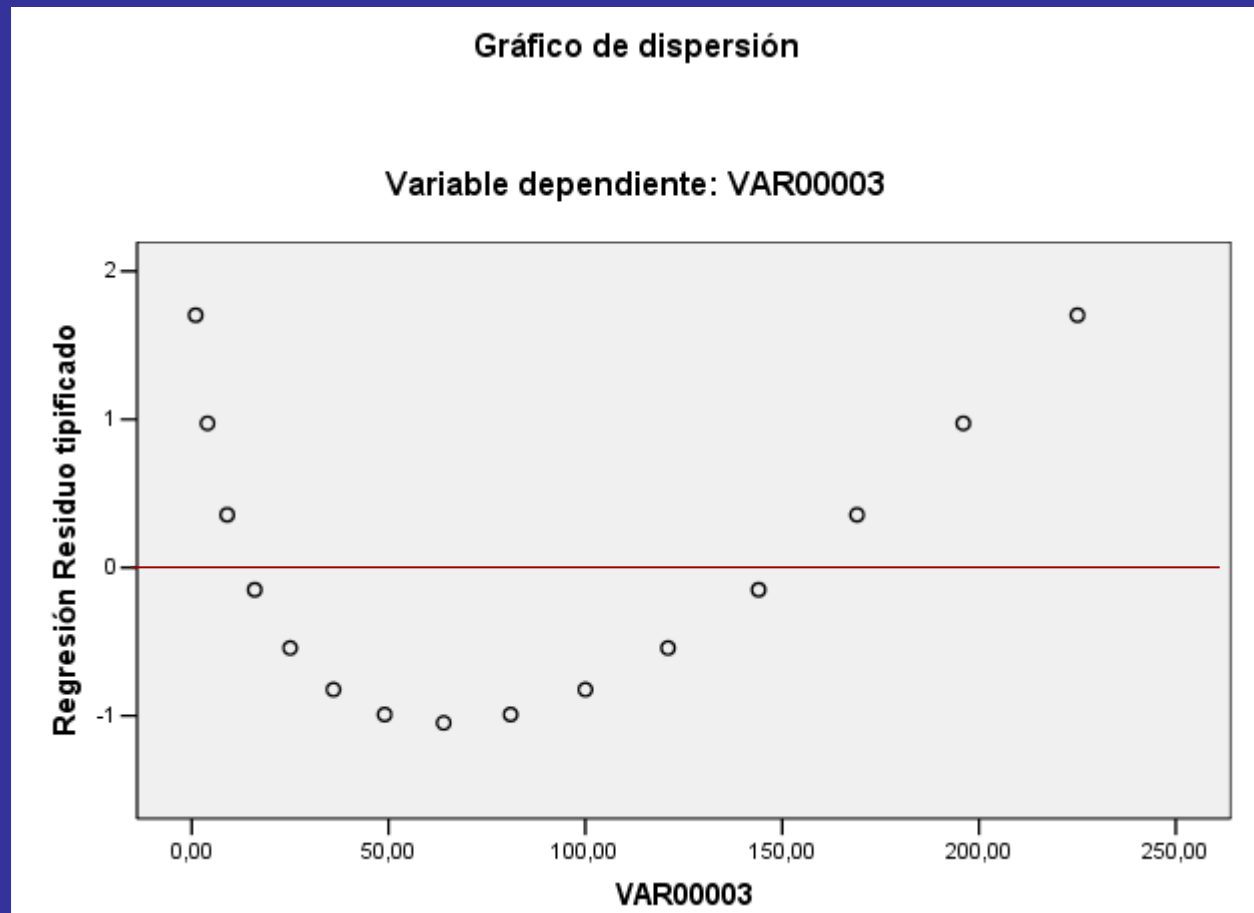
Análisis de Regresión Múltiple



1	1	3
1	2	5
1	3	7
1	4	9
1	5	11
1	6	13
1	7	15
1	8	17
1	9	19
1	10	21
1	11	23
1	12	25
1	13	27
1	14	29
1	15	



Análisis de Regresión Múltiple



Análisis de Regresión Múltiple

Histograma

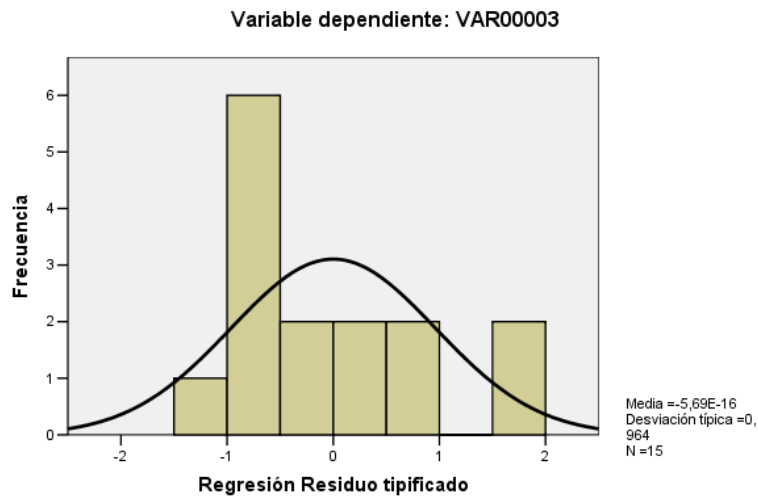
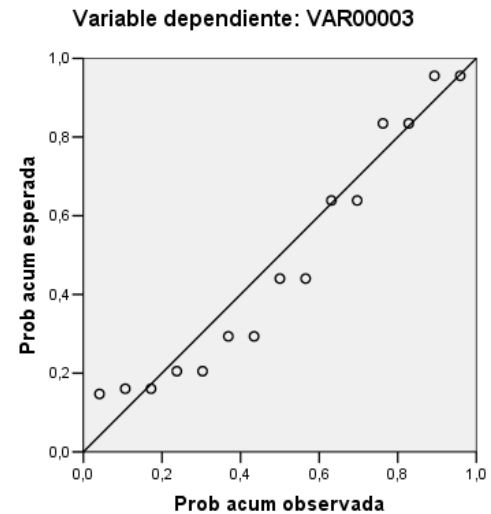
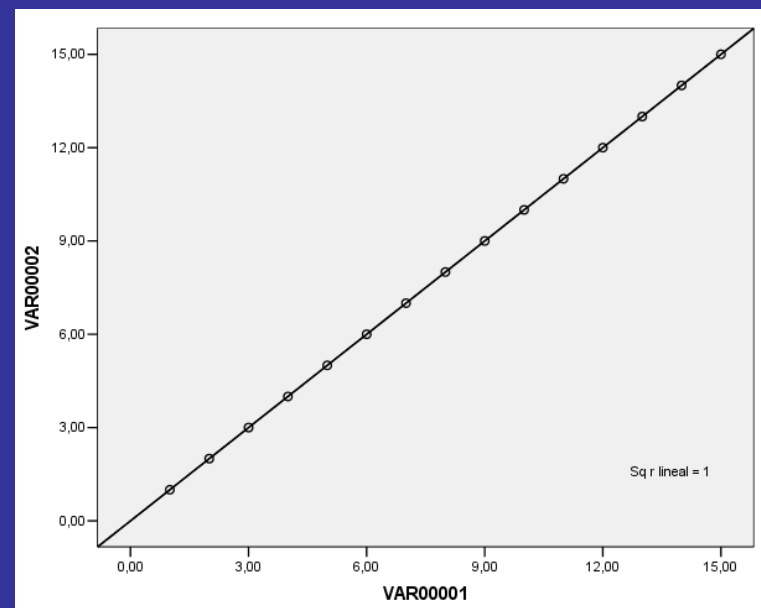
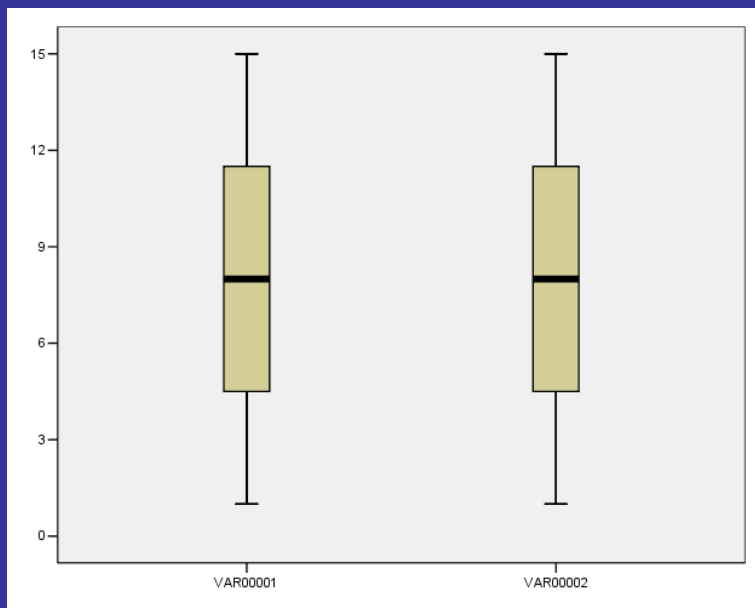


Gráfico P-P normal de regresión Residuo tipificado



Análisis de Regresión Múltiple

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15



Análisis de Regresión Múltiple

- Es una extensión de la regresión simple a más de una variable independiente:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_pX_p + e$$

a: valor medio estimado de Y cuando X es 0

b: cambio medio estimado en Y cuando X1 es 1, permaneciendo constantes el resto de independientes

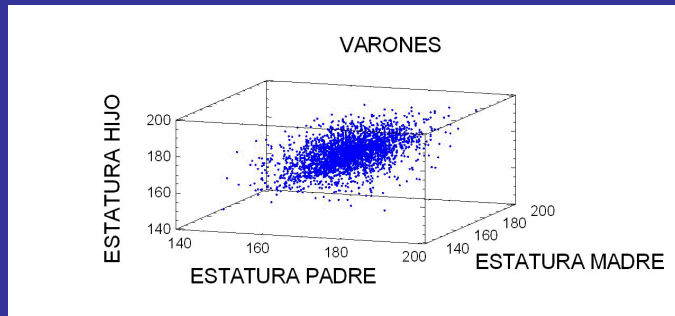
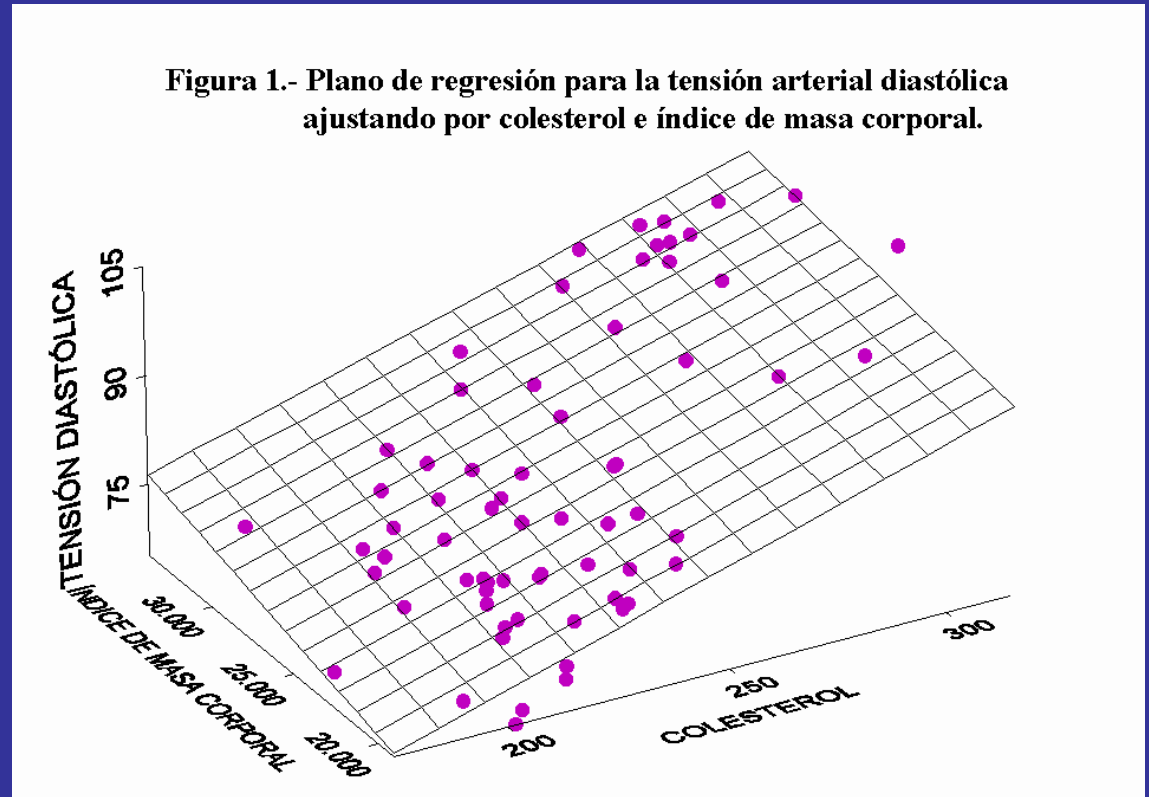
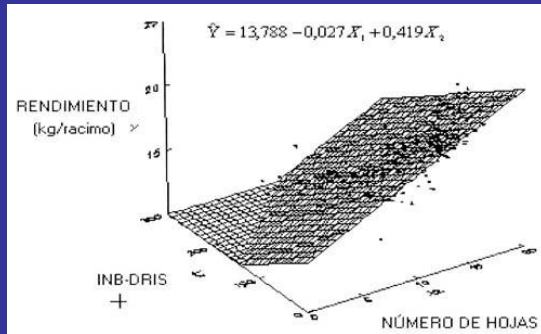
e: varios aspectos que no podemos cuantificar exactamente

- errores puramente aleatorios
- efectos de otras variables no incluidas en la ecuación
- errores de medición
- errores derivados de la forma funcional elegida

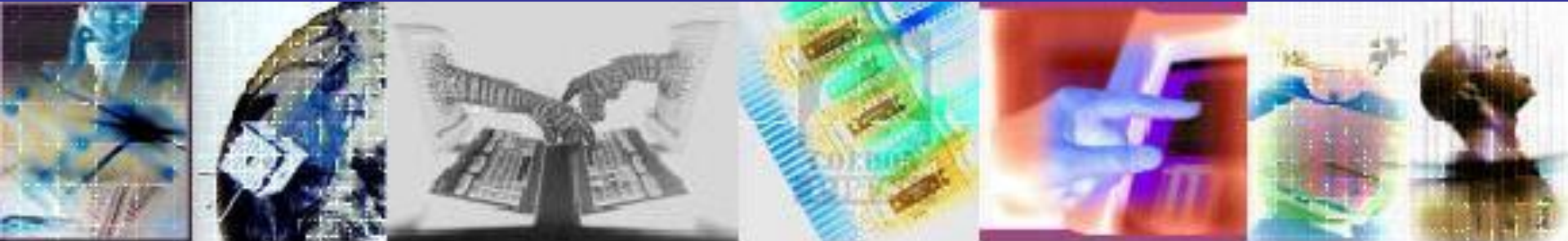
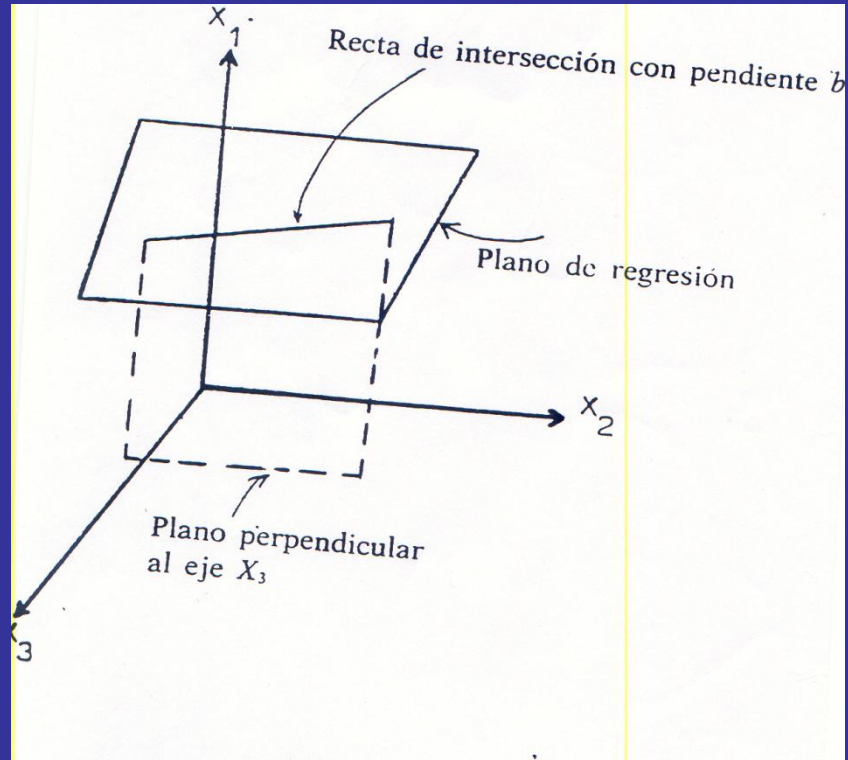


Análisis de Regresión Múltiple

En este caso ajustamos a un plano en lugar de a una recta

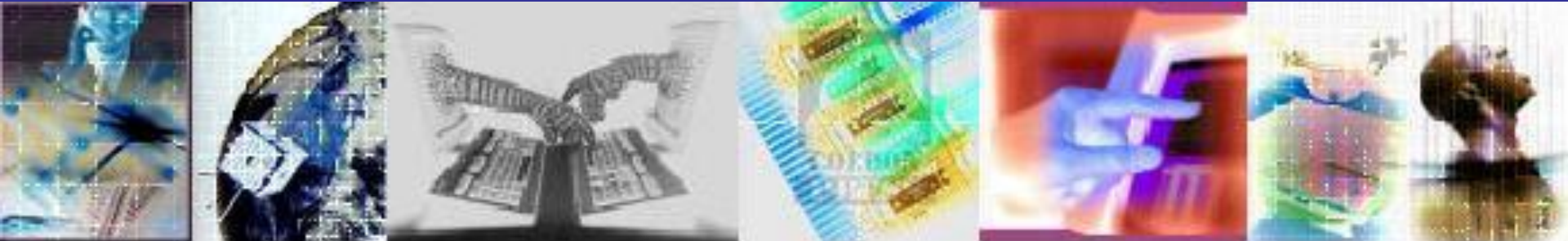


Análisis de Regresión Múltiple



Análisis de Regresión Múltiple

Definición: “La regresión múltiple no es más que una técnica estadística para especificar, estimar e interpretar un modelo explicativo en el que una variable dependiente se estudia en función de una serie de variables explicativas independientes”



Análisis de Regresión Múltiple

Supuestos paramétricos en el Análisis Multivariable

- **Normalidad Multivariable**
- **Homocedasticidad**
- **Tamaño muestral: $n > 30$**
- **Linealidad**
- **Nivel de medición continuo**



Análisis de Regresión Múltiple

TÉCNICAS DE DEPENDENCIA

Variable/s Dependiente/s	Variable/s Independiente/s	Técnica
1 Nivel de medición Continuo	1 Nivel de medición Continuo	REGRESIÓN SIMPLE
1 Nivel de medición Continuo	> 1 Nivel de medición Continuo	REGRESIÓN MÚLTIPLE
> 1 Nivel de medición Continuo	> 1 Nivel de medición Continuo	CORRELACIÓN CANÓNICA
1 Nivel de medición Nominal	> 1 Nivel de medición Continuo	ANÁLISIS DISCRIMINANTE
1 Nivel de medición Nominal (v. Dicotómica)	> 1 Nivel de medición Continuo	REGRESIÓN LOGÍSTICA
1 Nivel de medición Continuo	1 Nivel de medición Nominal	ANOVA Unidireccional
1 Nivel de medición Continuo	> 1 Nivel de medición Nominal	ANOVA Factorial o Bidireccional
> 1 Nivel de medición Continuo	1 Nivel de medición Nominal	MANOVA Unidireccional
> 1 Nivel de medición Continuo	1 Nivel de medición Nominal (v. Dicotómica)	T de Hotelling
> 1 Nivel de medición Continuo	> 1 Nivel de medición Nominal	MANOVA Factorial o Bidireccional

Análisis de Regresión Múltiple

1 PASOS DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- 1 Selección de las variables
- 2 Análisis de la matriz de correlaciones
- 3 Cuantificar la relación. Correlación múltiple
- 4 Cálculo de los coeficientes de correlación parcial
- 5 Significación de los estadísticos hallados
- 6 Obtener la ecuación de regresión e intervalo de confianza de la predicción



Análisis de Regresión Múltiple

Coeficiente de correlación múltiple

R: indica en qué medida el conjunto de las Independientes influye sobre la dependiente

R²: coeficiente de determinación múltiple

- Oscila entre 0 y 1
- R = 0 no indica ausencia de relación puede haberla pero no ser lineal
- Tiende a estar sobreestimado por ello se utiliza R ajustada



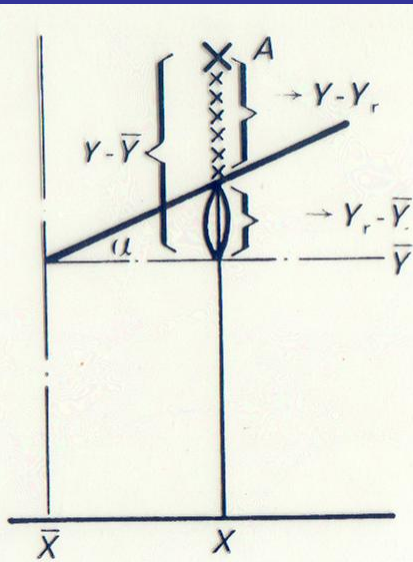
Análisis de Regresión Múltiple

Significación estadística de R múltiple

A partir de la F de Fisher

$$TSS = RSS + ESS$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y_i - \hat{Y})^2$$



$$R = \frac{RSS}{TSS}$$

$$F = \frac{RSS}{ESS}$$

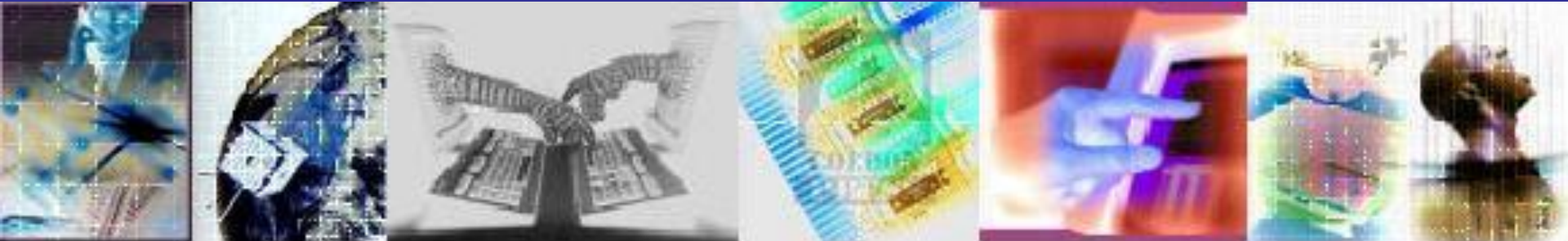


Análisis de Regresión Múltiple

Selección del número óptimo de predictores
Objetivo conseguir la máxima R múltiple con el menor número de variables (parsimonia informativa)

Regresión múltiple escalonada (SPSS)

- “Forward Selection”
- “Backward Selection”
- “Stepwise Selection”



Análisis de Regresión Múltiple

2 CONCEPTOS EN LA INTERPRETACIÓN DE RESULTADOS

Coeficientes “b” y Beta

b indica el aumento que experimentará la variable dependiente por cada unidad de incremento en la independiente manteniéndose constante el resto

Beta es el coeficiente de regresión estandarizado que permite la comparación del aporte de cada una de las independientes



Análisis de Regresión Múltiple

Constante o “intercept”

a es el valor de la constante y con el punto en el que el plano corta al eje de ordenadas

Tolerancia Este concepto está relacionado con La MULTICOLINEALIDAD. La tolerancia de una Variable indica en qué medida una variable Independiente está relacionada con las demás. Se refiere al % de variabilidad no explicada Por las otras variables:

Tolerancia = $1 - R^2_{ix}$ En el 1º paso $R^2_{ix} = 0$, $T = 1$



Análisis de Regresión Múltiple

VIF es otro indicador de multicolinealidad

Un valor de tolerancia reducido y elevados valores de **VIF** denotan una elevada Multicolinealidad

Durbin Watson indica si existe o no independencia entre las independientes. Si su valor tiende a 2 hay independencia si tiende a 0 o a 4 correlación, positiva y negativa respectivamente

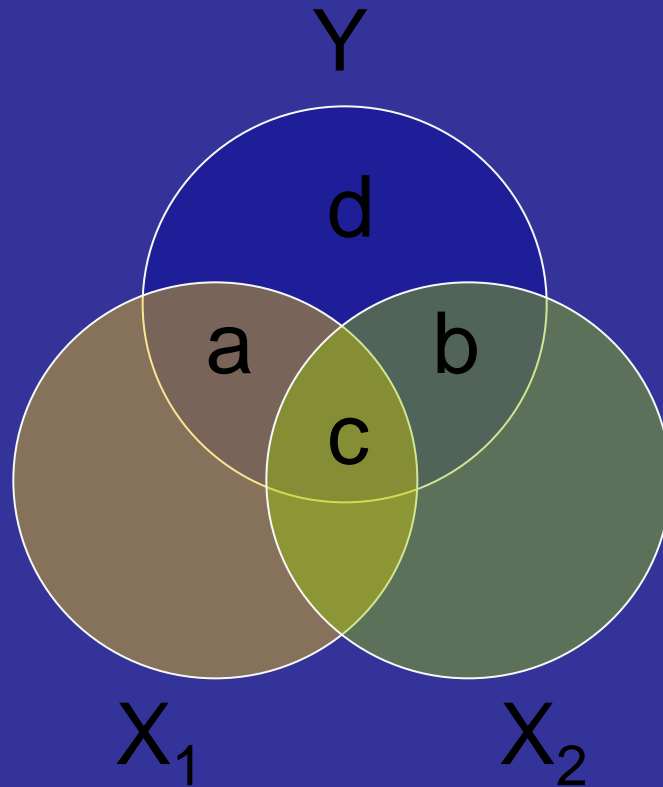


Análisis de Regresión Múltiple

El valor t indica si la relación de cada una de las variables independientes es significativa respecto a la dependiente, para ello debe de ser superior a 2,5



Análisis de Regresión Múltiple



- a** Varianza de Y explicada por X1
- b** Varianza de Y explicada por X2
- c** Varianza de Y explicada conjuntamente por X1 y X2
- d** Varianza de Y no explicada por X1 y X2



Análisis de Regresión Múltiple

Coeficientes de correlación parcial y semiparcial

Correlación parcial: es la correlación de una variable independiente X_i y una dependiente Y cuando se suprimen todos los efectos del resto de las independientes tanto sobre X_i como sobre Y

Correlación semiparcial: refleja la correlación entre una variable independiente X_i y una dependiente Y mientras se controlan los efectos predictivos de las otras independientes sobre X_i . Elevado al cuadrado sirve para observar la varianza única de cada independiente, del total de la varianza explicada por las independientes, que parte corresponde a cada independiente



Análisis de Regresión Múltiple

	Media	Desv. típ.	Mínimo	Máximo
SATISFACCIÓN	10.68	4.50	4	20
NIVEL COGNOSCITIVO	26.44	4.59	15	35
DEPRESIÓN	21.36	6.89	3	41
ANSIEDAD	13.99	3.51	0	18
SALUD	24.20	6.04	12	40
CAPACIDAD FUNCIONAL	16.62	25.03	0	100
APOYO SOCIAL	5.58	2.81	0	9



Análisis de Regresión Múltiple

Correlaciones

	Satisfacción	Nivel cognoscitivo	Depresión	Ansiedad	Salud	Capacidad funcional	Apoyo social
SATISFACCIÓN							
NIVEL COGNOSCITIVO	-.266**						
DEPRESIÓN	-.685**	.258**					
ANSIEDAD	-.677**	.188**	.545**				
SALUD	.731**	-.285**	-.578**	-.641**			
CAPACIDAD FUNCIONAL	.201**	-.282**	-.216**	-.172*	.358**		
APOYO SOCIAL	.410**	-.095	-.286**	-.292**	.289**	.135	
SEXO	.167*	-.230**	-.193**	-.268**	-.146*	.067	-.082

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

* La correlación es significativa al nivel 0,05 (bilateral).



Análisis de Regresión Múltiple

Coefficientes^a

Modelo	Coeficientes estandarizados		Coeficientes estandarizados	t	Sig.	Correlaciones			Estadísticos de colinealidad	
	B	Error típ.	Beta			Orden cero	Parcial	Semiparcial	Tolerancia	FIV
1 (Constante)	-2.438	.931	.	-2.620	.010					
SALUD	.543	.037	.731	14.557	.000	.731	.731	.731	1.000	1.000
2 (Constante)	7.100	1.557	.	4.560	.000					
SALUD	.374	.040	.504	9.263	.000	.731	.564	.412	.667	1.500
DEPRESIÓN	-.256	.035	-.393	-7.218	.000	-.684	-.470	-.321	.667	1.500
3 (Constante)	12.994	1.990	.	6.531	.000					
SALUD	.282	.044	.379	6.432	.000	.731	.429	.272	.516	1.940
DEPRESIÓN	-.213	.035	-.326	-6.046	.000	-.684	-.408	-.256	.615	1.625
ANSIEDAD	-.327	.074	-.255	-4.445	.000	-.675	-.312	-.188	.545	1.836
4 (Constante)	11.292	1.996	.	5.659	.000					
SALUD	.269	.043	.361	6.285	.000	.731	.422	.259	.511	1.955
DEPRESIÓN	-.199	.034	-.305	-5.770	.000	-.684	-.393	-.237	.607	1.648
ANSIEDAD	-.301	.072	-.234	-4.176	.000	-.675	-.296	-.172	.538	1.858
APOYO SOCIAL	.241	.070	.150	3.440	.001	.410	.247	.141	.885	1.130

^a Variable dependiente: SATISFACCIÓN.

Satisfacción = 11,9 + 0,2 salud – 0,1 depresión -0,3 ansiedad + 0,2 apoyosoc.
 Satisfacción = 0,36 salud – 0,3 depresión - 0,23 ansiedad + 0,15 apoyosoc



Análisis de Regresión Múltiple

Resumen del modelo^e

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	.731 ^a	.534	.531	3.08	
2	.798 ^b	.637	.633	2.73	
3	.820 ^c	.672	.667	2.60	
4	.832 ^d	.692	.685	2.52	1.742

^a Variables predictoras: (Constante), SALUD.

^b Variables predictoras: (Constante), SALUD, DEPRESIÓN.

^c Variables predictoras: (Constante), SALUD, DEPRESIÓN, ANSIEDAD.

^d Variables predictoras: (Constante), SALUD, DEPRESIÓN, ANSIEDAD, APOYO SOCIAL.

^e Variable dependiente: SATISFACCIÓN.

No son incluidas: “Nivel cognoscitivo”,
“Capacidad funcional”
“Sexo”



Análisis de Regresión Múltiple

Variables excluidas^e

Modelo	Beta dentro	t	Sig.	Correlación parcial	Estadísticos de colinealidad	
					Tolerancia	
1	SEXO	-.067 ^a	1.116	.266	.082	.979
	NIVEL COGNOSCITIVO	-.061 ^a	-1.175	.242	-.086	.920
	DEPRESIÓN	-.393 ^a	-7.218	.000	-.470	.667
	ANSIEDAD	-.351 ^a	-5.834	.000	-.395	.590
	CAPACIDAD FUNCIONAL	-.067	-1.250	.213	-.092	.870
	APOYO SOCIAL	.217 ^a	4.342	.000	.305	.917
2	SEXO	.014 ^b	.310	.757	.023	.962
	NIVEL COGNOSCITIVO	-.022 ^b	-.472	.638	-.035	.907
	ANSIEDAD	-.255 ^b	-4.445	.000	-.312	.545
	CAPACIDAD FUNCIONAL	-.072 ^b	-1.516	.131	-.111	.870
	APOYO SOCIAL	.170 ^b	3.749	.000	.267	.895
	3	SEXO	-.024 ^c	-.554	.581	-.041
NIVEL COGNOSCITIVO		-.028 ^c	-.633	.528	-.047	.906
CAPACIDAD FUNCIONAL		-.056 ^c	-1.223	.223	-.090	.864
APOYO SOCIAL		.150	3.440	.001	.247	.885
4	SEXO	.002 ^d	.046	.963	.003	.896
	NIVEL COGNOSCITIVO	-.028 ^d	-.654	.514	-.049	.906
	CAPACIDAD FUNCIONAL	-.062 ^d	-1.407	.161	-.104	.863

^a Variables predictoras en el modelo: (Constante), SALUD.

^b Variables predictoras en el modelo: (Constante), SALUD, DEPRESIÓN.

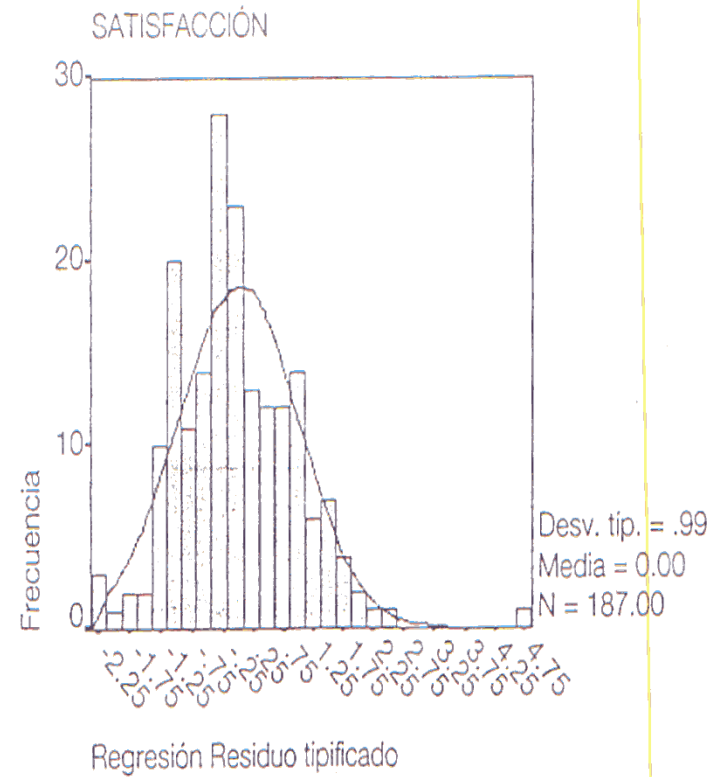
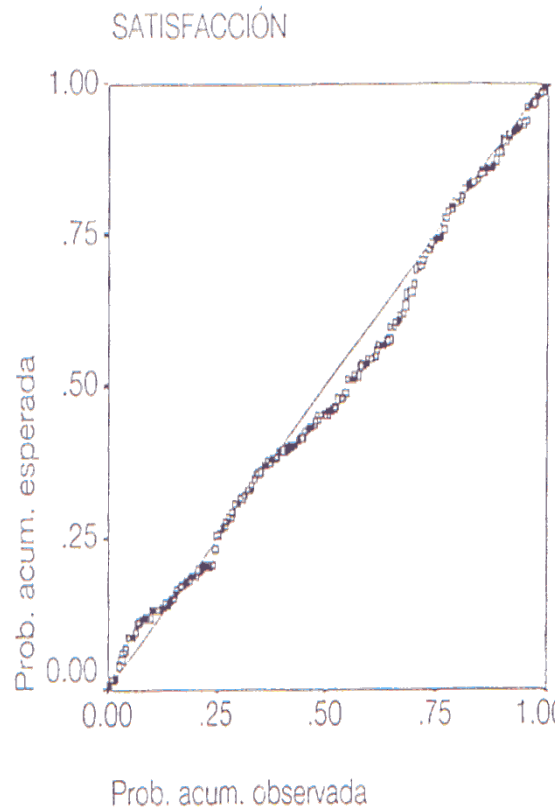
^c Variables predictoras en el modelo: (Constante), SALUD, DEPRESIÓN, ANSIEDAD.

^d Variables predictoras en el modelo: (Constante), SALUD, DEPRESIÓN, ANSIEDAD, APOYO SOCIAL.

^e Variable dependiente: SATISFACCIÓN.



Análisis de Regresión Múltiple



Análisis de Regresión Múltiple

Categoría	Variable Casado	Variable Soltero	Variable Viudo
Persona Casada	1	0	0
Persona Soltera	0	1	0
Persona Viuda	0	0	1
Persona Divorciada	0	0	0



Análisis de Regresión Múltiple

Satisfacción = a + b₁ Salud + b₂ casado + b₃ soltero + b₄ viudo

Satisfacción = 20 + 3 Salud + 11,5 casado + 5 soltero - 3 viudo

Para los **casados** que tienen asignado el valor 1 en la variable CASADO y 0 en el resto al sustituir estos valores en la ecuación:

Satisfacción = 20 + 3 Salud + 11,5 casado = 31,5 + 3 salud

Para los **solteros**:

Satisfacción = 20 + 3 Salud + 5 soltero = 25 + 3 salud

Para los **viudos**:

Satisfacción = 20 + 3 Salud - 3 viudo = 17 + 3 salud

Para los **divorciados**:

Satisfacción = 20 + 3 Salud

