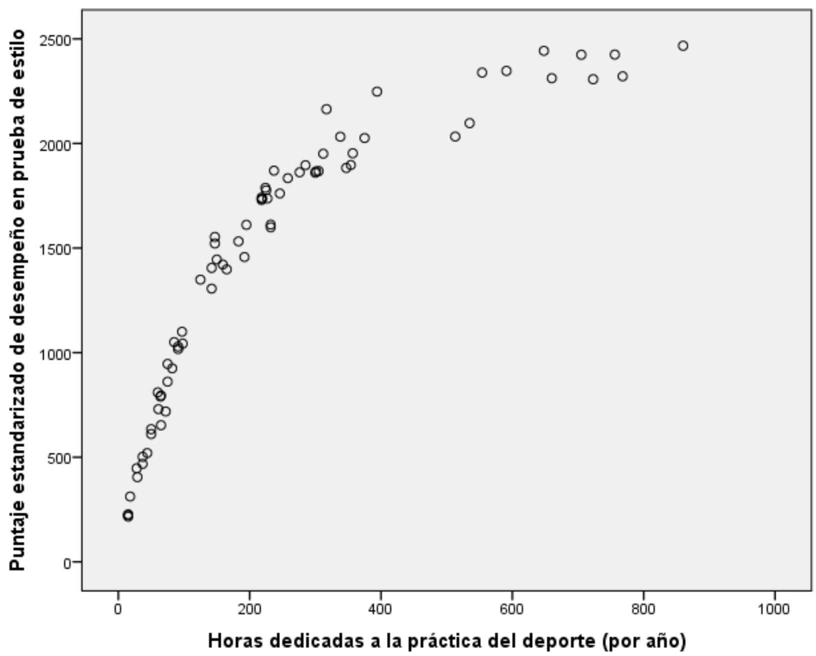
Análisis de regresión no lineal

Muchas veces, a pesar de que existe una relación entre la variable independiente y la dependiente, esta no sigue una relación lineal. Imaginemos un estudio de desempeño en deportistas (Y) y horas de práctica (X_1), rápidamente se podría pensar que mayor tiempo de práctica equivaldría siempre a un incremento en el desempeño, pero sabemos también, sin embargo que, a partir de cierto nivel de práctica, uno alcanza una meseta de rendimiento. Como puede observarse en el gráfico, parece existir una relación de proporcionalidad para las primeras 200 horas de práctica a al año, pero a partir de ese número empieza a notarse un cambio en la tendencia. Es posible que si formulemos un modelo lineal con estas variables obtuviésemos resultados significativos, sin embargo nuestro Coeficiente de determinación (R^2) no sería óptimo. En este caso es conveniente aplicar una regresión No lineal.

La interpretación de los parámetros (β, ε, R, R^2, etc) es exactamente igual que en la regresión lineal



Existen varios tipos de regresiones no lineales, pero nos concentraremos en dos modelos en particular:

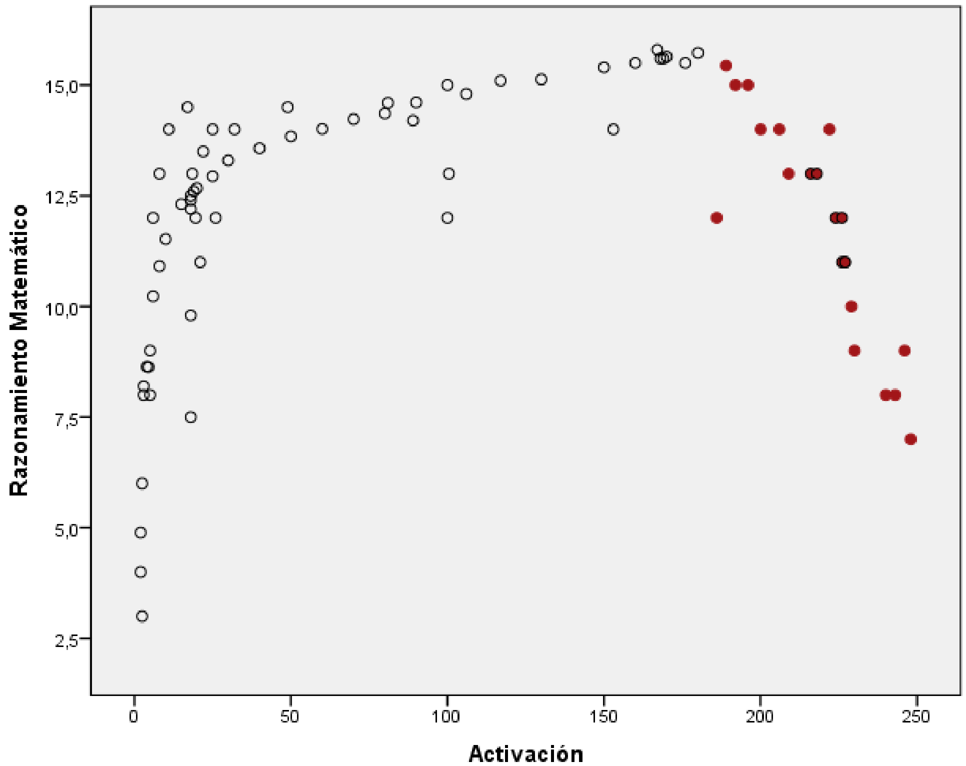
Cuadrático:

Este modelo es similar al modelo de una regresión lineal simple, sólo que incorpora una variable independiente adicional, ésta es la elevación al cuadrado de la primera variable independiente. Se tiene así el siguiente modelo:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \pm \varepsilon$$

El ejemplo de inicio del capítulo sigue una lógica cuadrática. Otro ejemplo similar es el siguiente, un grupo de psicólogos está interesado en estudiar la relación que existe entre la activación fisiológica (X) y el desempeño en una prueba de matemática (Y). Ellos encuentran que conforme se incrementa la activación también lo hace el rendimiento en la prueba, sin embargo esto sólo funciona hasta un punto, pasado ese nivel de activación el desempeño empieza a disminuir rápidamente (nerviosismo, bloqueo cognitivo, etc).

Para la primera parte se espera una relación directa ($\beta_1 > 0$), mientras que para la segunda se espera una asociación inversa ($\beta_2 < 0$) A continuación se presenta el diagrama de dispersión que ellos encontraron.



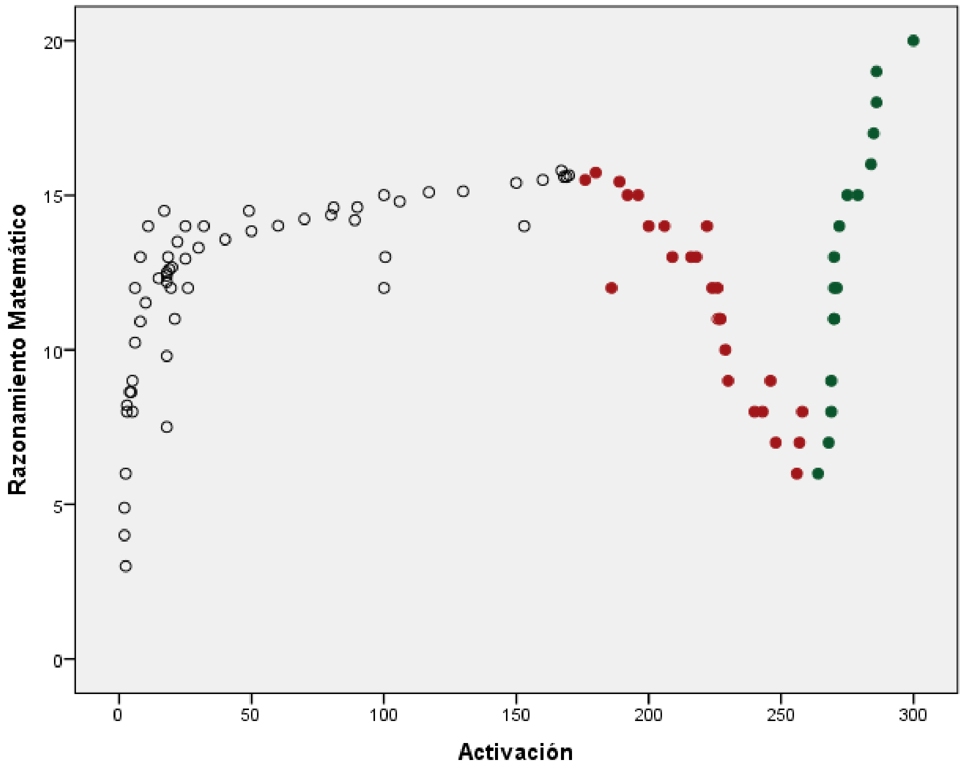
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \pm \varepsilon$$

Polinomial:

Al igual que el modelo cuadrático, los modelos polinomiales se construyen sobre la base del modelo lineal. En este caso se tiene un polinomio de más de dos grados (la potencia a la que se eleva), por ejemplo, un modelo a la tercera potencia (cúbico) o un modelo a la cuarta potencia. En general siguen la siguiente fórmula:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_p X^p \pm \varepsilon$$

Por ejemplo, imaginemos que el estudio del ejemplo anterior (activación fisiológica (X) y el desempeño en una prueba de matemática (Y)) obtiene unos resultados ligeramente diferentes: Durante la investigación se encontraron datos extremos que contradecían la teoría y el modelo formulado, por lo que fueron descartados como outliers, sin embargo una revisión posterior llegó a la conclusión de que estos datos podrían ser registrados en un modelo cúbico. El equipo explica que en una situación normal un nivel de activación demasiado elevado iba acompañado de un desempeño bajo en matemáticas, pero que pasado cierto nivel de activación, los puntajes volvían a incrementarse hasta llegar a niveles de desempeño muy elevados. Al igual que en el caso anterior se esperaba en la primera parte una relación directa ($\beta_1 > 0$) y para la segunda una asociación inversa ($\beta_2 < 0$), pero además se incorpora una tercera tendencia, nuevamente positiva ($\beta_3 < 0$) A continuación se presenta el diagrama de dispersión que ellos encontraron:



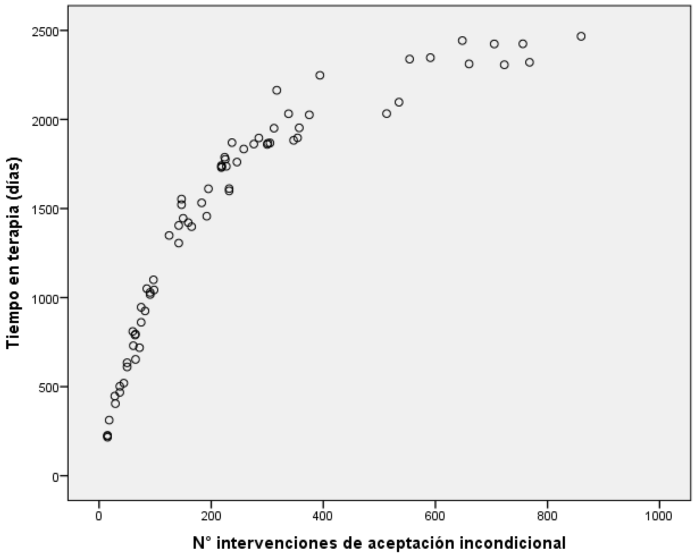
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \pm \varepsilon$$

Ejercicio de Regresión no lineal cuadrática y cúbica (Reg. No lineal.sav)

A una investigadora le interesa conocer los factores que influyen en el tiempo que una persona asiste a tratamiento psicoterapéutico. La investigadora, adherente al Enfoque Centrado en la Persona de Carl Rogers, cree que la cantidad de veces que un(a) psicoterapeuta dé muestras de aceptación incondicional a su cliente influirá en que el paciente permanezca más tiempo en psicoterapia. Así, luego de contactar a 30 psicoterapeutas y de que estos accedieran a grabar una sesión, decide contar el número de veces que cada terapeuta hace una intervención en la que demuestre su aceptación incondicional por la persona en consulta.

La investigadora tiene como hipótesis que las muestras de aceptación tendrán una relación directa en el tiempo que la persona permanezca en terapia.

a) Realice un diagrama de dispersión en SPSS e interprete el gráfico. ¿Se cumple la hipótesis de la investigadora? ¿Cuántas tendencias aparecen en el gráfico?



En un comienzo, parecería haber evidencia gráfica que sustente la hipótesis de proporcionalidad directa entre ambas variables. Sin embargo, luego comienza a verse cierto estancamiento y, probablemente, a descender. Habría entonces dos tendencias en la gráfica.

b) Cuando tenemos dos tendencias opuestas y sucesivas en un mismo gráfico, nos encontramos ante un caso de regresión no lineal, los datos deben ser representado en una ecuación cuadrática. Los **modelos cuadráticos** son representados de la siguiente forma:

$Y = \beta 0 + \beta 1X + \beta 2X2 + \varepsilon$.

Esta ecuación representa que, cuando X toma valores mayores, X2 domina la tendencia; ello quiere decir que X2 lleva los puntos hacia arriba, si el coeficiente $\beta 2$ es positivo, o hacia abajo, si el coeficiente $\beta 2$ es negativo.

Según lo observado en el gráfico, ¿cuál de los siguientes modelos sería el que mejor presenta lo observado?

i. <u>Modelo 1:</u> $Y = \beta 0 + \beta 1X + \beta 2X2 + \varepsilon$ *con* $\beta 2 > 0$
ii. <u>Modelo 2:</u> $Y = \beta 0 + \beta 1X + \beta 2X2 + \varepsilon$ *con* $\beta 2 < 0$

Podríamos hipotetizar que el modelo 2 sería el más adecuado. Sin embargo, para confirmarlo habría falta el análisis estadístico.

c) Intentando entender los resultados, la investigadora se planteó la posibilidad de que, a partir de cierto momento, las muestras de aceptación incondicional dejan de surtir efecto debido a que el terapeuta podría no necesariamente sentir una aceptación incondicional por su cliente. En tal sentido, la relación entre el tiempo que una persona permanece en terapia y las manifestaciones de aceptación condicional podrían ser **no lineales**.

Para hacer una regresión con un modelo cuadrático es necesario calcular la variable X2 de la siguiente manera:

Transformar ➔Calcular variable ➔ Variable de destino: X2 ➔ Expresión numérica: X2 ➔ Aceptar.**

Luego, realice en SPSS una regresión lineal múltiple con X y X2 como V.I. Examine el ajuste del modelo, su significancia y, luego, los contrastes unilaterales. ¿El modelo representa bien los datos? ¿Es mejor que el modelo que sólo incluye X?

<u>Modelo lineal</u>

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Tiempo
  /METHOD=ENTER Aceptación.
```

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,872[a]	,761	,757	321,938

a. Variables predictoras: (Constante), N° intervenciones de aceptación incondicional

ANOVA[a]

Modelo		Suma de cuadrados	Gl	Media cuadrática	F	Sig.
1	Regresión	22718048,592	1	22718048,592	219,193	,000[b]
	Residual	7151442,282	69	103644,091		
	Total	29869490,873	70			

a. Variable dependiente: Tiempo en terapia (días)

b. Variables predictoras: (Constante), N° intervenciones de aceptación incondicional

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	801,333	58,339		13,736	,000
	N° intervenciones de aceptación incondicional	2,682	,181	,872	14,805	,000

a. Variable dependiente: Tiempo en terapia (días)

<u>Modelo cuadrático</u>

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Tiempo
  /METHOD=ENTER Aceptación X2.
```

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,975[a]	,951	,949	147,436

a. Variables predictoras: (Constante), X2, N° intervenciones de aceptación incondicional

El modelo representa los datos en .98, lo cual es bastante alto.
Asimismo, las VI explican el 95,1% de la variación de los puntajes en la VD.
Finalmente, el modelo cuadrático (R2C = .95) es mejor que el lineal (R2C = .76).

ANOVA[a]

Modelo		Suma de cuadrados	Gl	Media cuadrática	F	Sig.
1	Regresión	28391355,719	2	14195677,859	653,057	,000[b]
	Residual	1478135,154	68	21737,282		
	Total	29869490,873	70			

a. Variable dependiente: Tiempo en terapia (días)
b. Variables predictoras: (Constante), X2, N° intervenciones de aceptación incondicional

El modelo es significativo (p = .00 < .05)

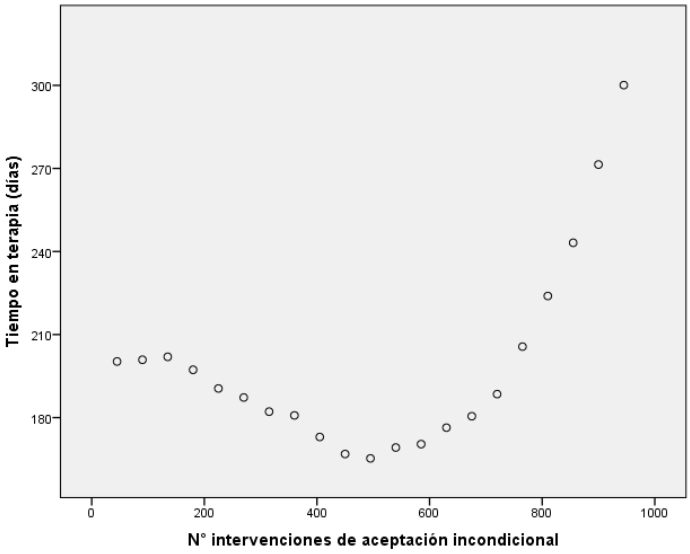
Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	355,456	38,413		9,254	,000
	N° intervenciones de aceptación incondicional	6,979	,279	2,269	25,051	,000
	X2	-,006	,000	-1,463	-16,155	,000

a. Variable dependiente: Tiempo en terapia (días)

Podemos concluir que la tendencia predominante en este modelo es negativa.

d) Al concluir la investigación, el investigador deseaba conocer si ocurre lo mismo con otra muestra de psicoterapeutas y clientes. Para ello contactó a 21 nuevos psicoterapeutas y repitió el proceso original. Describa los resultados con un gráfico de puntos. Para ello, utilice la base de datos **Reg. cúbica**.



Vemos 3 tendencias en el gráfico: al comienzo los puntajes van en aumento, pero rápidamente decaen. Luego remontan y siguen subiendo y continúan incrementándose.

e) Al observar el gráfico, el psicólogo plantea que la coherencia por parte del terapeuta (la expresión de los sentimientos que surgen en la relación terapéutica) generaría que, con el tiempo, pueda nuevamente incrementar las intervenciones que demuestran una aceptación incondicional. En ese sentido, la data recolectada pareciera seguir un **modelo cúbico** representado como $Y = \beta 0 + \beta 1X + \beta 2X2 + \beta 3X3 + \varepsilon$, siendo $\beta 1 > 0$, $\beta 2 < 0$ y $\beta 3 > 0$. Genere las variables X2 y X3 y emita una opinión sobre cuál es el modelo estadísticamente más relevante empleando el SPSS.

Modelo lineal

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Tiempo
  /METHOD=ENTER Aceptación.
```

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,513[a]	,263	,224	30,793

a. Variables predictoras: (Constante), N° intervenciones de aceptación incondicional

Modelo cuadrático

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Tiempo
  /METHOD=ENTER Aceptación X2.
```

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,949[a]	,901	,890	11,613

a. Variables predictoras: (Constante), X2, N° intervenciones de aceptación incondicional

Modelo cúbico

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Tiempo
  /METHOD=ENTER Aceptación X2 X3.
```

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,999[a]	,997	,997	1,966

a. Variables predictoras: (Constante), X3, N° intervenciones de aceptación incondicional, X2

El modelo cúbico (R2C = .997) es el que mejor representa los datos