

# Segundo modelo para una epidemia:

## Modelo Monomolecular

- Se remonta al menos a 1909 (Mitscherlich) pero no fue usado para epidemias
- Apropiado en general para epidemias monocíclicas
- Muy diferente al modelo exponencial
- $dy/dt$  no es proporcional a  $y$ 
  - Razón: el inóculo producido por individuos infectados en el curso de la epidemia no produce infecciones adicionales (durante esa epidemia)
- Sin embargo,  $dy/dt$  es proporcional a la “intensidad” de enfermedad libre
- En este caso representamos la intensidad de enfermedad sobre una escala de proporción
  - Por lo tanto, la intensidad de libre enfermedad o nivel es  $(1-y)$

# Modelo Monomolecular

- $dy/dt = \underline{r_M}(1-y)$

- $r_M$  es el parámetro de tasa (unidades 1/tiempo)

- Más específicamente, el parámetro es igual a  $\phi Q$  el producto de:  
 $r_M = \phi \cdot Q$

- La cantidad de inóculo ( $Q$ ) disponible para la infección de plantas libres de enfermedad (u otra unidad bajo estudio ej. Hojas, raíces, etc) y

- La probabilidad por unidad de tiempo ( $\phi$ ) que una unidad de inóculo cause infección (es una tasa)

- Más específicamente  $\phi$  es el número individuos nuevos que se han enfermado por individuos sanos. por unidad de inóculo por unidad de tiempo

- Es una función del ambiente, susceptibilidad, y “eficiencia” de inóculo (función de la agresividad del patógeno)

# Modelo Monomolecular

- $dy/dt = r_M(1-y) = r_M - r_M y$
- Al inicio de la infección,  $y$  es baja (por definición)
  - Por lo tanto,  $1-y = 1$  ó  $r_M y = 0$
  - Y,  $dy/dt = r_M$  (una constante) – acumulación de \$ a **interés simple**
- La integración de la forma diferencial, lleva al modelo no lineal

$$y = 1 - (1-y_0) \exp(-r_M t)$$

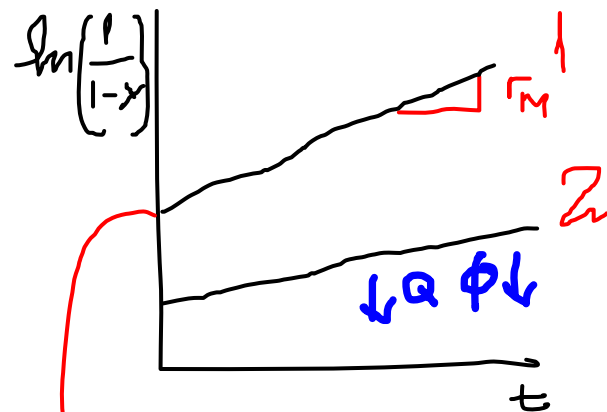
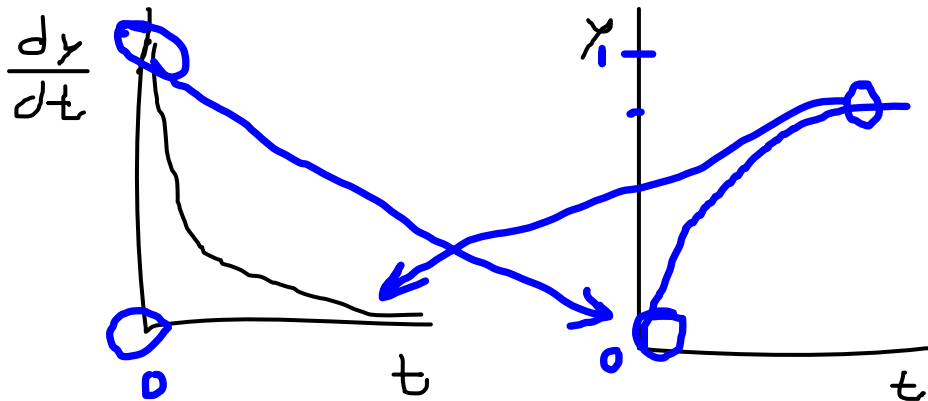
- Como siempre,  $y_0$  es la intensidad inicial de enfermedad (*constante de integración*)
- En su forma linealizada

$$\ln(1/(1-y)) = \ln(1/(1-y_0)) + r_M t \quad \text{ó} \quad y^* = y_0^* + r_M t$$

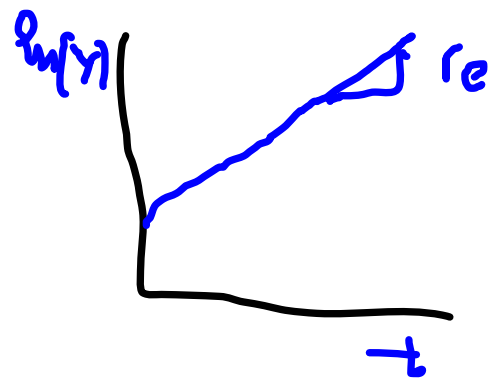
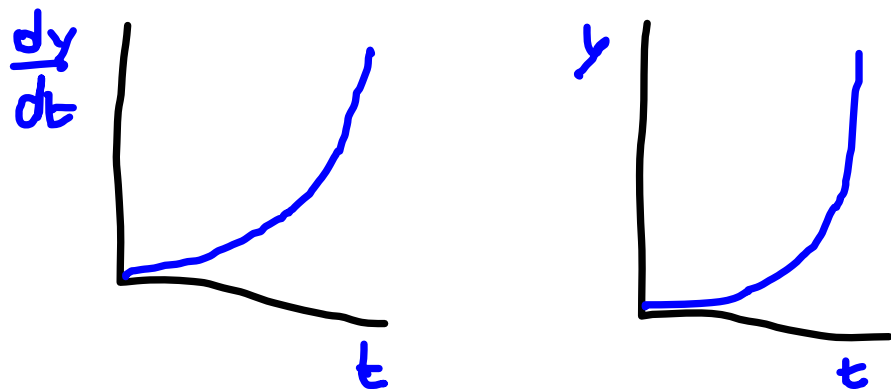
Donde  $y^* = \ln(1/(1-y))$  es conocida como la

**Transformación de Infección Múltiple**

M



E



# Nota

- A un valor bajo de  $y$

$$y = 1 - (1 - y_0) \exp(-r_M t)$$

Simplificando:

$$y \approx y_0 + r_M t$$

$$y^*, \ln(1/(1-y)) \approx y$$

$y$

$y^*$

*Ejemplo:*  $y=0.02$  (2%),  $\ln(1/(1-0.02)) \approx 0.0202$

Pero, hay una diferencia a un valor mayor de  $y$

$$y = 0.50 \text{ (50\%)} , \ln(1/(1-0.50)) = 0.693 > 0.50 \text{ (+15.3\%)}$$

# Tercer modelo para una epidemia:

## Modelo Logístico

- Se remonta al menos a 1838 (Verhulst) (crecimiento de poblaciones)
- Conocido también como autocatalítico (reacciones químicas)
- Apropriado para epidemias **policíclicas**
- Es una generalización del exponencial
- El incremento de la tasa absoluta de enfermedades proporcional a ambos:

• **y:**

(intensidad de enfermedad), más enfermedad significa más inóculo, y más enfermedad y así sucesivamente; y

• **1-y:**

(sana), más enfermedad significa menos plantas sanas (o menos tejido sano) disponible para la infección, por lo tanto la tasa de crecimiento decrece (hasta 0)

•  $dy/dt = r_L y(1-y)$   **$r_L$  = parámetro de la tasa (1/tiempo)**

## Modelo Logístico $dy/dt = r_L y(1-y)$

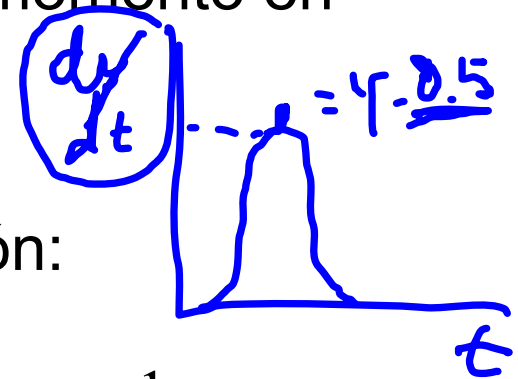
El parámetro ( $r_L$ ) como en el modelo exponencial:

- Producción de “inóculo” por individuos infectados
  - (Relativo a la **infectividad** de los individuos enfermos)
- Probabilidad que el “inóculo” alcance un huésped libre de enfermedad desde un individuo infectado
- Probabilidad que el “inóculo” en contacto con un huésped libre de enfermedad cause infección
- Tiempo que demora un individuo recientemente infectado a producir “inóculo”
- El tiempo que un individuo infectado esté produciendo inóculo”

$$r_L = f(a, b, c, d, e, \dots)$$

# Modelo Logístico $dy/dt = r_L y(1-y)$

$dy/dt$  alcanza su valor máximo cuando al momento en  $y=0.5$



La integración lleva a la siguiente expresión:

$$y = \frac{1}{1 + \exp\left(-\left(\ln\left(\frac{y_0}{1-y_0}\right) + r_L * t\right)\right)} \quad \text{ó} \quad y = \frac{1}{1 + \left(\frac{1-y_0}{y_0}\right)e^{-r_L t}}$$

- existen otras fórmulas

## Nota:

- $y_0$  es la constante de integración y es la intensidad inicial de enfermedad (y a  $t=0$ )
- Es un modelo de dos parámetros



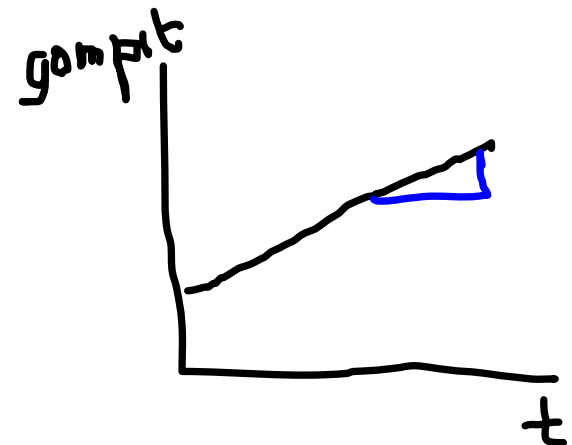
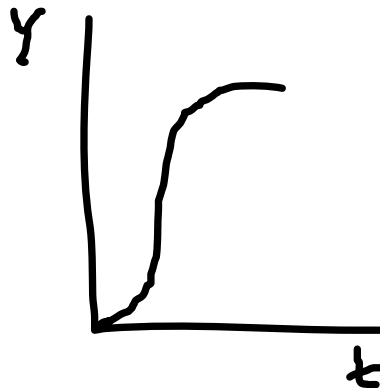
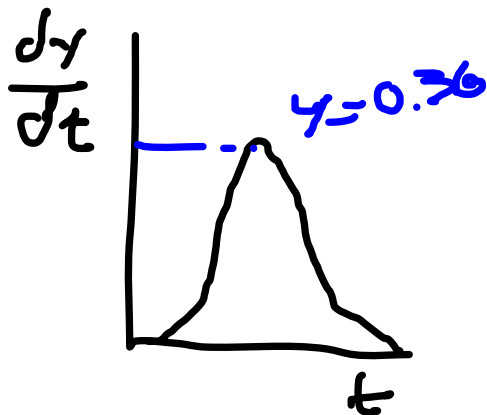
# Modelo Gompertz

Forma diferencial:  $dy/dt = r_g y [-\ln(y)]$

Integrada:  $y = e^{[-Be^{-r_g t}]}$

$\text{gompfit}(x)$

Transformada:  $-\ln[-\ln(y)] = -\ln[-\ln(y_0)] + r_g t$



## Otros modelos: Richards

- Richards (1959) crecimiento en animales

$$dy/dt = [r_R / (\underline{\eta} - 1)] y (1 - y^{\underline{\eta}})$$

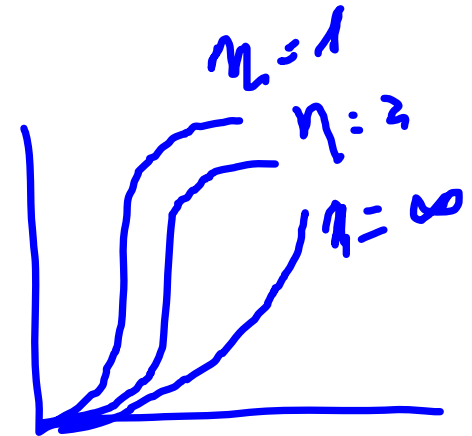
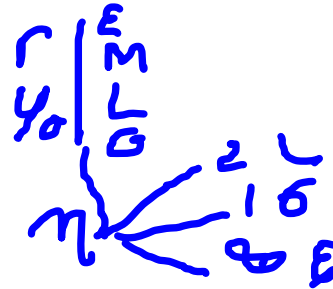
- $\eta$  es un parámetro afectado a la forma de la curva

- $f(1-y) = 1 - y^{\eta-1}$

- El modelo incluye a todos los otros

$\eta =$

- 2 logística
- 1 Gompertz ( $\eta \rightarrow \infty$ )
- 0 Monomolecular
- $\rightarrow \infty$  Exponencial



- Al fin, se considera una familia de modelos para epidemias; la forma específica depende de:

$f(y)$ , con  $f(y)=y$  en todos los casos



- Existen otros modelos posibles de este tipo

- $\frac{dy}{dt}=r \cdot f(y)f(1-y)$

- Propiedades:

- Un parámetro de “tasa”
  - Una variable aleatoria (y) modelada en términos de su cambio absoluto respecto al tiempo t
  - Luego de su integración y algo de álgebra, puede ser escrita como:
- En muchos casos, esta ecuación es lineal en todos los parámetro
- Excepción; para Richards,  $y^*$  involucra un parámetro con  $\eta < 1$

$$\ln\left(\frac{1}{1-y^{1-\eta}}\right) = \left(\frac{1}{1-y_0^{1-\eta}}\right) + r_R t$$

## Otros modelos: Weibul

- Desarrollado por Fisher & Tipper (1928) y derivado independientemente por Weibul (1939)
- Se ha usado en estudio de sobrevivencia pero también ha sido utilizado en modelos de crecimiento, de dosis/respuesta y modelo de progreso de enfermedades (Pennypacker et al. 1980)

$$\frac{dy}{dt} = \frac{c}{b} \left( \frac{t-a}{b} \right)^{c-1} \exp \left[ - \left( \frac{t-a}{b} \right)^c \right]$$

Handwritten annotations: A red circle around the parameter  $c$ , a blue underline under the parameter  $b$ , a blue arrow pointing from the exponent  $c$  to a handwritten  $\gamma$ , and a handwritten  $\gamma_0$  to the right.

- en los cuales  $a$ ,  $b$  y  $c$  son parámetros
  - $a$  = representa un parámetro de ubicación y es la ocurrencia más temprana de la enfermedad
  - $b$  = es un parámetro de escala inversamente proporcional a la tasa de incremento de la enfermedad (unidades de tiempo)
  - $c$  = parámetro sin unidad y controla la forma de la curva de la tasa ( $dy/dt$  vs.  $t$ ) y el punto de inflexión

# Otros modelos: Weibul

- Por integración se obtiene

$$y = 1 - \exp \left[ - \left( \frac{t - a}{b} \right)^c \right]$$

- Cuando  $c =$

- 1 monomolecular con  $r_M = 1/b$  ;  $B = \exp(a/b)$

- Hay un punto de inflexión cuando  $c$  es  $> 1$  el cual desplaza a valores mayores de  $y$  cuando  $c$  aumenta

- El punto de inflexión ocurre a  $y = 1 - \exp[(1/c) - 1]$  ;  $t = b[(c-1)/c]^{1/c} + a$

- De la ecuación anterior se determina que un punto de inflexión  $y = 0.5$  ocurre para un valor de  $c = 3.26$  análogo al modelo logístico
- Dubey (1967) demostró que la curva es simétrica cuando  $c = 3.6$
- De ahí, los modelos Weibul y logístico no son idénticos, pero Pennypacker demostró que pueden ser similares en un amplio rango de condiciones

## Otros modelos: Weibul

- La forma linealizada puede ser expresada como:

$$\left[ \ln \left( \frac{1}{1-y} \right) \right]^{1/c} = \frac{-a}{b} + \frac{t}{b}$$

con la intercepción de  $-a/b$  y una pendiente de  $1/b$

- o de otra forma

$$\ln \left[ \ln \left( \frac{1}{1-y} \right) \right] = -c \ln(b) + c \ln(t - a)$$

con una intercepción de  $-c \ln(b)$  y una pendiente  $c$

*Dado que la primera ecuación linealizada tiene uno de los parámetros en el lado izquierdo, no puede ser utilizada para estimar los tres parámetros. En este caso, se usa la segunda opción, pero debe conocerse exactamente el valor de a, o sea, el momento del inicio de la epidemia*

# Eligiendo el modelo

- Conocimiento del sistema huésped-patógeno
- Sin embargo, más de un modelo puede ser apropiado, en particular para epidemias policíclicas
- Una combinación de **métodos gráficos y estadísticos** son de gran valor

## • Gráficos

### • **Graficar $y$ versus $t$**

- Es bueno para separar monomolecular de otros
- Dificultad para separar lo otros sin muchos datos (sobre un rango de  $y$ )

### • **Graficar $dy/dt$ versus $t$**

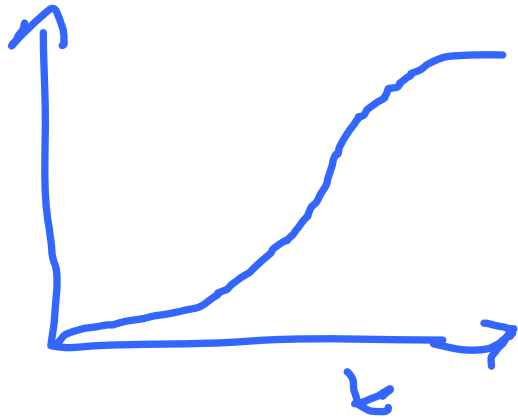
- Muy sensitivo para elegir entre varios modelos (especialmente monomolecular y otros)
- Aún así, requiere un **amplio rango** de valores de  $y$



# Eligiendo el modelo

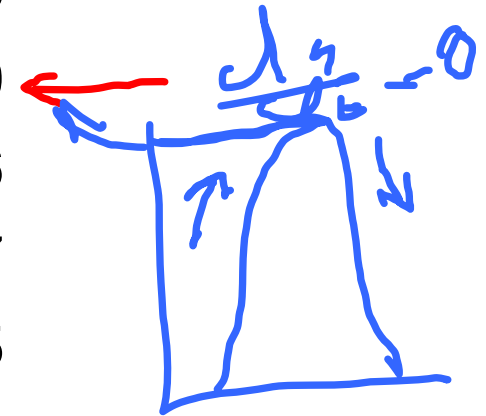
**No se puede observar  $dy/dt$  por lo tanto debe ser estimado**

$t$	$y$	$\Delta y$	$\Delta y / \Delta t$	$\frac{dy}{dt}$
0	0,010			
10	0,050	0,040	0,004	
20	0,150	0,100	0,010	
30	0,400	0,250	0,025	
40	0,700	0,300	0,030	
50	0,860	0,160	0,016	
60	0,930	0,070	0,007	
70	0,980	0,050	0,005	
80	0,990	0,010	0,001	



$1-y$   
 $\downarrow$   
 $y \rightarrow 1; y=1$

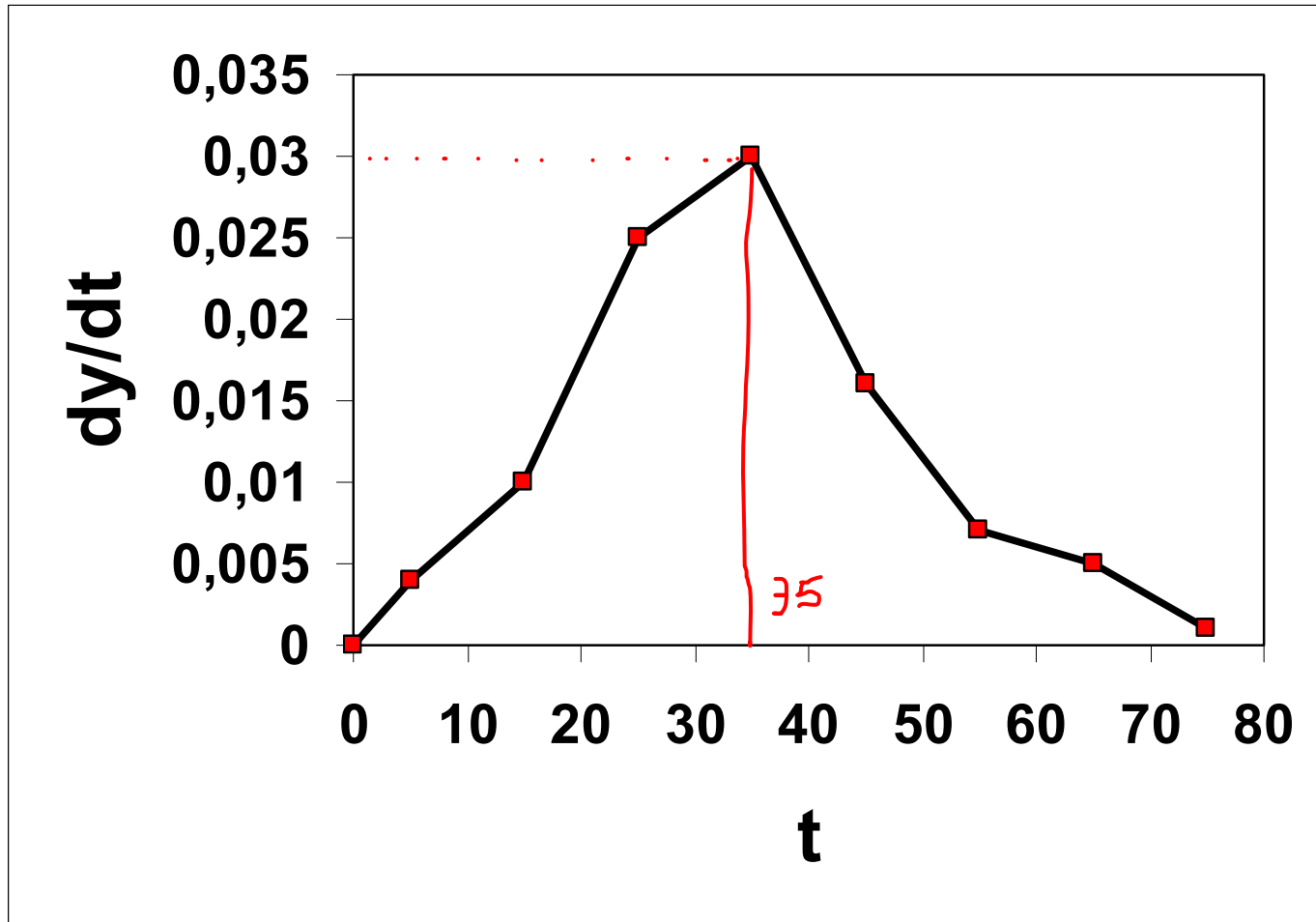
$1-x \rightarrow 0$



**Cada tasa se calcula entre dos puntos, por lo tanto en valor de abscisa usado es el punto medio de cada par de tiempos, ej.: entre el para 0-10 el punto medio es 5**

	$t$	$y$	$\Delta y$	$\Delta y / \Delta t$	$\leftarrow \frac{dy}{dt}$
5	0	0,010			
15	10	0,050	0,040	0,004	
25	20	0,150	0,100	0,010	
35	30	0,400	0,250	0,025	
45	40	0,700	0,300	0,030	$y = 0,55$
55	50	0,860	0,160	0,016	
65	60	0,930	0,070	0,007	
75	70	0,980	0,050	0,005	
	80	0,990	0,010	0,001	

Usar el punto medio entre pares de datos de t correspondientes a y para cada  $dy/dt$



# Eligiendo el modelo

## •Gráfico

- Graficar  $y$  versus  $t$
- Graficar  $dy/dt$  versus  $t$
- Graficar  $y^*$  versus  $t$ 
  - Esto implica varios gráficos
    - $\ln(y)$ ,  $\ln [1/(1-y)]$ , logit, gompit
  - Encontrar uno que dé la **línea más recta**

- Considerar:
  - Mono(0)
  - Rich ( $1/2$ )
  - Gomp ( $\sim 1$ )
  - Rich (1.5)
  - logistic (2)
  - Rich (3)
  - [incrementando  $\eta$ ]

## •Estadístico

- Regresión por cuadrados mínimos ordinarios

$$Z = b_0 + b_1X + e$$

$$y^* = y_0^* + r.t + e$$

- Una regresión separada por cada modelo posible
- Considerar  $R^2$ , CME y residuales ( $e$  versus  $t$ )

## •Regresión por cuadrados mínimos ordinarios

$$Z = b_0 + b_1X + e$$

$$y^* = y_0^* + r.t + e$$

- El mejor ajuste corresponde a más alto  $R^2$ , menor CME (la línea más recta)
- Sin embargo,  $y^*$  es diferente para cada modelo y toda las estadística está basada en

$$\sum (y_j^* - \hat{y}_j^*)^2$$

- Por lo tanto, hay que tomar algunas precauciones
- La estadísticas no son para argumentar acuerdo entre  $y$  e  $\hat{y}$

- **Aproximación estadística**

- Regresión por cuadrados mínimos

- **Regresión no lineal de cuadrados mínimos**

- Uso directo de  $y$  como función de  $t$  (y parámetros)
    - Más difícil
    - Uso de los residuales para un modelo ajustado dado
    - O ajustar un modelo flexible como Richards o Weibul
    - SON MUCHO MÁS COMPLICADOS (ESTADÍSTICAMENTE)
    - PERO SE PUEDE ESTIMAR DIRECTAMENTE  $\eta$
    - NO BUSCAN UNA LÍNEA RECTA

- **Cómo usar los resultados gráficos y estadísticos?**

- Algunas veces se desea el mejor ajuste posible
    - Pero esto puede significar un  $\eta$  diferente para cada epidemia
    - A menudo, es más satisfactorio el uso de un modelo simple que ajuste adecuadamente todas las curvas de enfermedades de interés

# **Análisis temporal de epidemias (continua)**

***Ahora vamos a analizar un ejemplo de datos y ver las salida del programa SAS en mayor detalle.***

***Se podrá ver las diferencias entre un modelo lineal y uno no lineal***

```

data a;
input t y;
datalines;
0 0.01
10 0.05
20 0.15
30 0.40
40 0.70
50 0.86
60 0.93
70 0.98
80 0.99
;
proc nlin data=a; title 'Logistic';
parameter y0star=.5 r=0.10;
bounds r>0;
model y=1/(1+exp(-(y0star+r*t))); — Logistica
output out=outLL r=r p=p;
proc gplot data=outLL;
plot r*t;
plot y*p0/overlay;
symbol1 w=1 h2 c=black V=circle l=none l=3;
symbol2 W=2 H=2 c=red V=triangle l=none l=1;
run;


```

*y < L       $\hat{p} + t$*



## Ajuste no lineal

Hay concordancia entre y e y predicha



Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Regression	2	4.2182	2.1091	7938.12	<.0001
Residual $= e$	7	<u>0.00186</u>	0.000266		
Uncorrected Total	9	<u>4.2201</u>			
Corrected Total	8	1.3640			

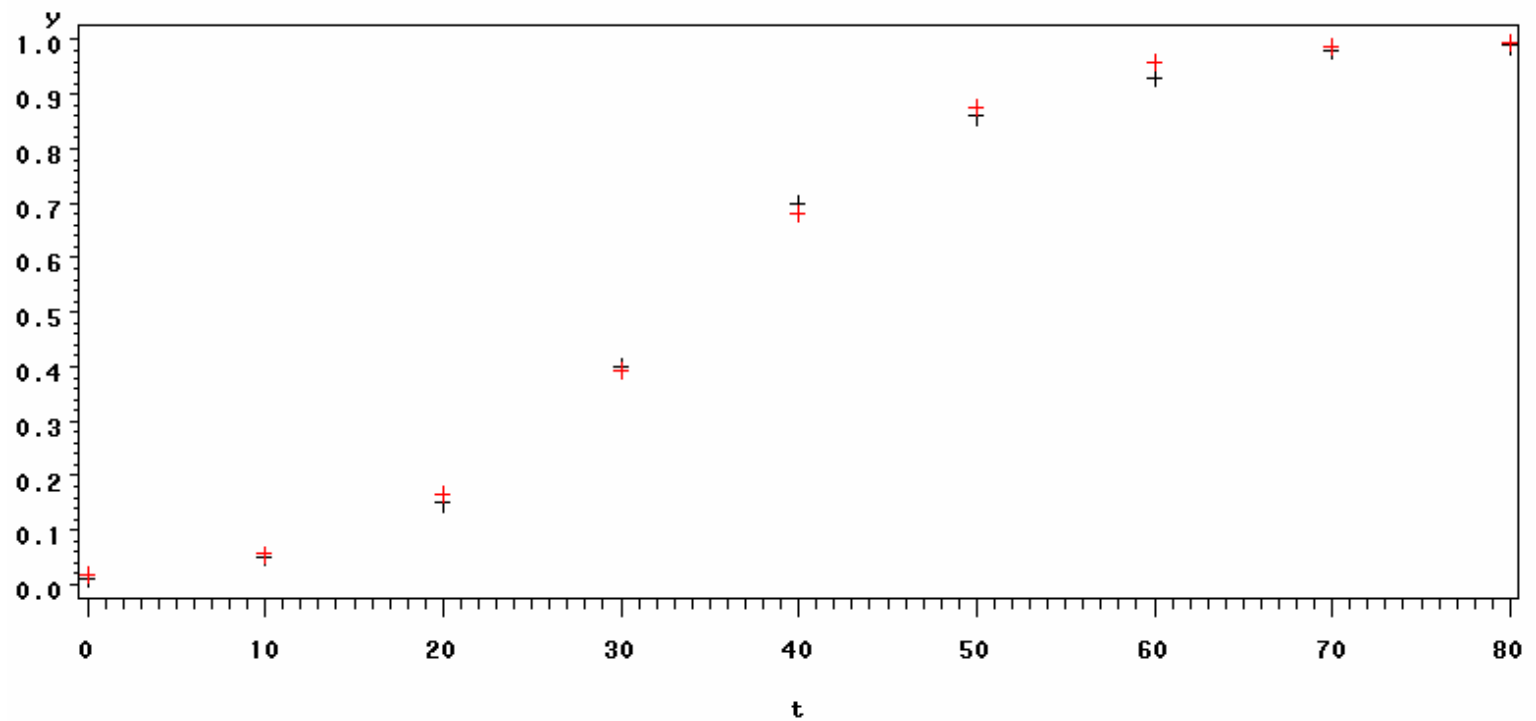
$$R^2 = 1 - \left( \frac{SCE}{SCT} \right)$$

Parameter	Estimate	Std Error	Approximate 95% Confidence Limits	
y0star	<u>-4.0026</u>	0.1600	-4.3810	-3.6242
r	0.1191	0.00458	0.1082	0.1299

$$\hat{y} = -4.0026 + 0.1191x$$

## Ajuste no lineal

Logistic



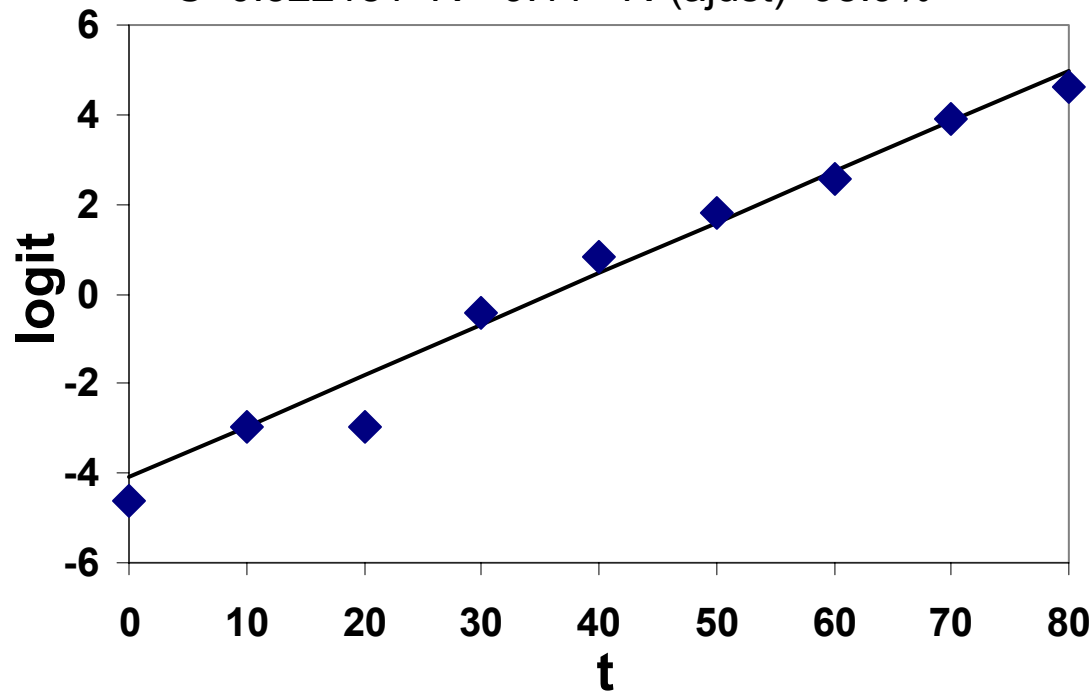
## Ajuste lineal

El acuerdo es entre *logit* y *logit* predichos, y no entre *y* e *y* predichas

### Gráfico de regresión

$$\text{logit} = -4.09147 + 0.113555 * t$$

S=0.322181 R<sup>2</sup>=9.11 R<sup>2</sup>(ajust)=98.9%



Los parámetros estimados pueden no ser los mismos entre ajuste lineal y no lineal

## Algunos comentarios acerca del análisis de la regresión

- En el caso de la regresión lineal, el parámetro de intercepción es intensidad de enfermedad transformada a  $t=0$  (no es  $y$  a  $t=0$ )
  - Ejemplo: estimación de  $\ln(y_0)$  para el modelo exponencial
- Además, los valores predichos a cada tiempo son de los valores de  $y^*$  transformados y no de  $y$ 
  - Aunque es conveniente para el análisis y cálculos, uno puede tener más interés en valores de  $y$  predichos
- Se puede usar la **de transformación** (inversa de la función) para obtener valores estimados de  $y_0$  y valores predichos de  $y$ 
  - Ejemplo: para el modelo exponencial  $y_0 = \exp(y_0^*)$
  - La de transformación es justo la reversa de la transformación original

## Algunos comentarios acerca del análisis de la regresión

- Se puede usar los valores de transformación para calcular la bondad adicional del ajuste
- Se puede determinar los intervalos de confianza para los estimadores de los parámetros
  - Para la intercepción del análisis de la regresión lineal, se puede determinar los intervalos de confianza directamente para  $y_0^*$ , pero No para  $y_0$
  - Para obtener los intervalos de confianza para  $y_0$ , se usa la de transformación de los límites del intervalo para  $y_0^*$

## Algunos comentarios acerca del análisis de la regresión

- Los *residuales* pueden ser usados para probar el modelo (diagnóstico del modelo) en el estilo de *modelaje empírico* para evaluar:
  - *Esta siendo usado el modelo corregido*
  - *Se han alcanzado los supuestos estadísticos*  
*Especialmente importante:*
    - *Normalidad de los errores*
    - *Independencia de los errores*
    - *Varianza constante de los errores*