

# Stata para usuarios de SPSS

**Aurelio Tobías**

Hospital de la Santa Creu i Sant Pau  
Barcelona, 21/01/2019

## 5. Modelos de regresión

## Modelos de regresión

- Respuesta continua
  - Regresión Lineal  
 $\mu = \alpha + \beta X$
- Respuesta binaria
  - Regresión Logística  
 $\log(\pi/(1-\pi)) = \alpha + \beta X$
- Respuesta discreta
  - Regresión de Poisson  
 $\log(\lambda) = \alpha + \beta X$

¿Qué tienen en común  
estos modelos de regresión?

## Modelos de regresión

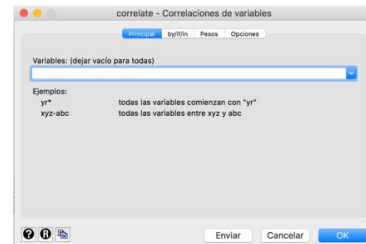
- Respuesta continua
  - Regresión Lineal, con  $X=(0,1)$   
 $\mu = \alpha + \beta X$       donde  $\alpha = \mu_0$  y  $\beta = \mu_1 - \mu_0$
- Respuesta binaria
  - Regresión Logística, con  $X=(0,1)$   
 $\log(\pi/(1-\pi)) = \alpha + \beta X$       donde  $\exp(\alpha) = \pi_0/(1-\pi_0)$  y  $\exp(\beta) = OR$
- Respuesta discreta
  - Regresión de Poisson, con  $X=(0,1)$   
 $\log(\lambda) = \alpha + \beta X$       donde  $\exp(\alpha) = \lambda_0$  y  $\exp(\beta) = IRR$

# Correlación

- Coeficiente de correlación de Pearson (y covarianza)

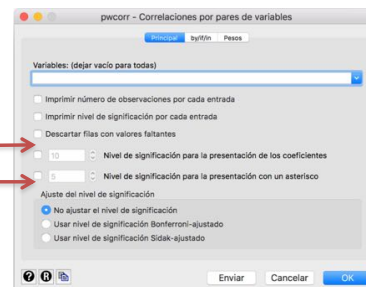
> Estadísticas > Sumarios, tablas y tests estadísticos > Sumarios y estadísticas descriptivas > Correlaciones y covarianzas

. **correlate vars**



> Estadísticas > Sumarios, tablas y tests estadísticos > Sumarios y estadísticas descriptivas > Correlaciones por pares

. **pwcorr vars, sig star(#)**



```
. use episer, clear
```

```
. correlate bone weight height
(obs=400)
```

	bone	weight	height
bone	1.0000		
weight	0.3954	1.0000	
height	0.5117	0.5117	1.0000

```
. pwcorr bone weight height , sig
```

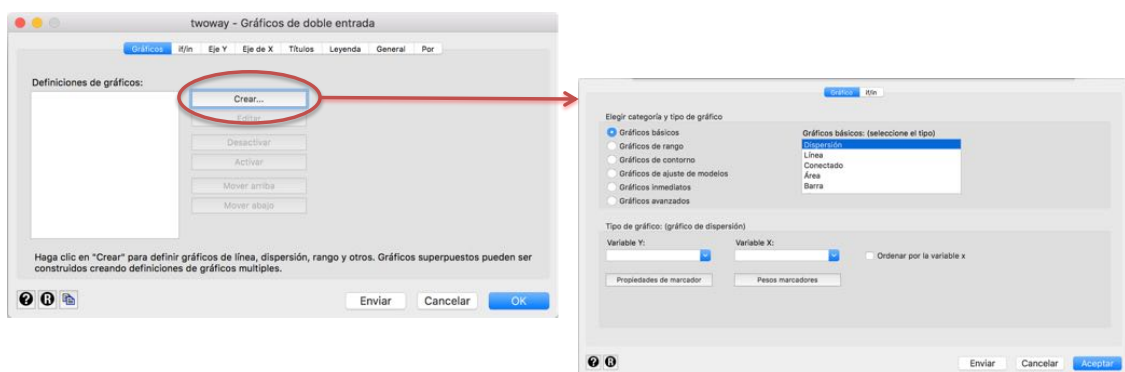
	bone	weight	height
bone	1.0000		
weight	0.3954 0.0000	1.0000	
height	0.5117 0.0000	0.5117 0.0000	1.0000

# Gráficos de dos dimensiones

- **Stata** permite superponer múltiples tipos de gráficos para mostrar cada relación bivariente
- Se tiene que definir cada tipo de gráfico (scatter, line, linear fit, etc.)
  - > Gráficos > Gráficos de doble entrada (dispersión, línea, etc.)
  - . **twoway type vary varx**
  - . **twoway (type vary varx) (type vary varx) ...**
- Hay una gran cantidad de opciones relacionadas con el tipo de trazado, eje, leyenda y títulos, que se pueden usar mediante la sintaxis y el menú de gráficos de dos vías, o mediante el Editor de gráficos (¡pero aquí los cambios no se guardarán como sintaxis!)

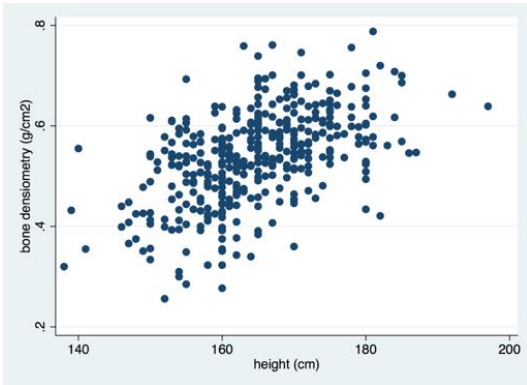
7

# Gráficos de dos dimensiones

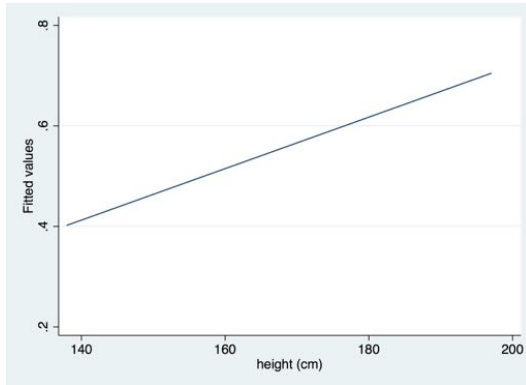


8

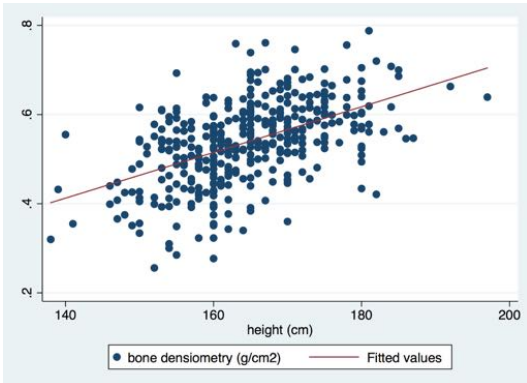
```
. twoway scatter bone height
```



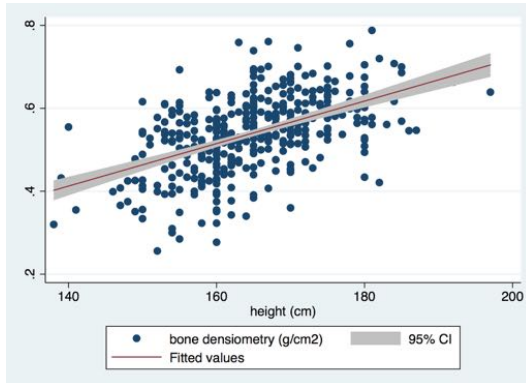
```
. twoway lfit bone height
```



```
. twoway (scatter bone height) (lfit bone height)
```



```
. twoway (scatter bone height) (lfitci bone height)
```



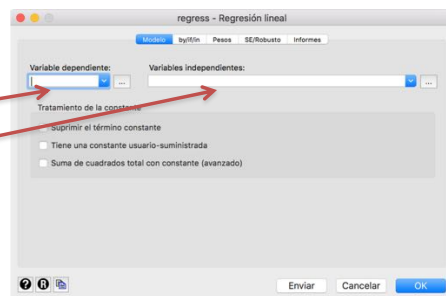
## Modelos de regresión

- Regresión lineal

- > Estadísticas > Modelos lineales y afines

- > Regresión lineal

- . regress vary varx

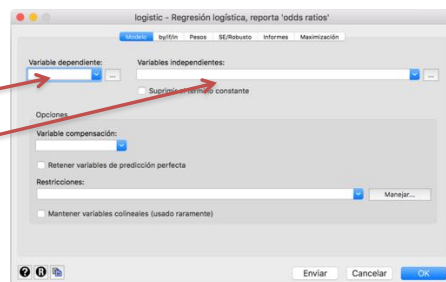


- Regresión logística

- > Estadísticas > Respuestas binarias >

- Regresión logística

- . logistic vary varx



```
. regress bone height
```

Source	SS	df	MS	Number of obs	=	400
Model	.878434619	1	.878434619	F(1, 398)	=	141.18
Residual	2.4763671	398	.006222028	Prob > F	=	0.0000
				R-squared	=	0.2618
				Adj R-squared	=	0.2600
Total	3.35480172	399	.008408024	Root MSE	=	.07888

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
height	.0051226	.0004311	11.88	0.000	.0042751 .0059702
_cons	-.3046196	.0710653	-4.29	0.000	-.4443298 -.1649093

```
. summarize height
```

Variable	Obs	Mean	Std. Dev.	Min	Max
height	400	164.5825	9.159582	138	197

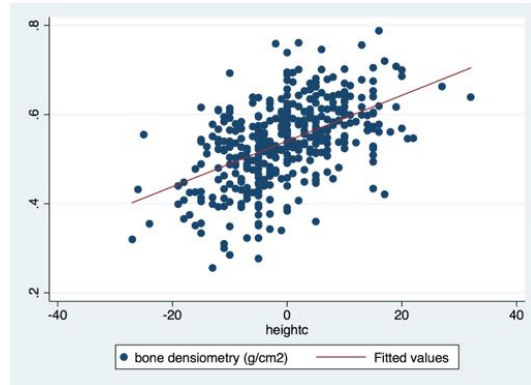
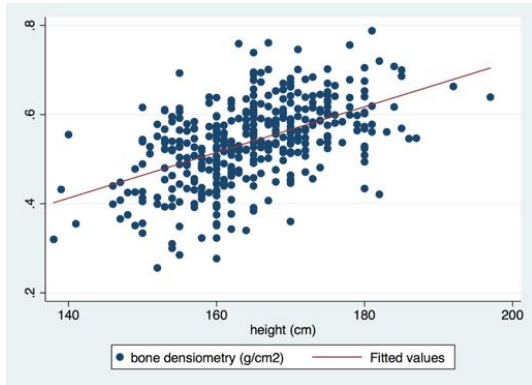
```
. generate heightc = height-165
```

```
. regress bone heightc
```

Source	SS	df	MS	Number of obs	=	400
Model	.878434619	1	.878434619	F(1, 398)	=	141.18
Residual	2.4763671	398	.006222028	Prob > F	=	0.0000
				R-squared	=	0.2618
				Adj R-squared	=	0.2600
Total	3.35480172	399	.008408024	Root MSE	=	.07888

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
heightc	.0051226	.0004311	11.88	0.000	.0042751 .0059702
_cons	.5406137	.0039481	136.93	0.000	.532852 .5483754

```
. twoway (scatter bone height) (lfit bone height) . twoway (scatter bone heightc) (lfit bone heightc)
```



```
. logistic art age
```

```
Logistic regression                               Number of obs   =       400
                                                    LR chi2(1)      =      102.19
                                                    Prob > chi2     =       0.0000
Log likelihood = -144.83762                       Pseudo R2      =       0.2608
```

	art	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age		1.094428	.0119299	8.28	0.000	1.071293 1.118061
_cons		.0020283	.0013107	-9.60	0.000	.0005716 .0071973

Note: \_cons estimates baseline odds.

```
. summarize age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	400	46.2	16.77538	19	93

```
. generate agec = age-45
```

```
. logistic art agec
```

```
Logistic regression                               Number of obs   =       400
                                                    LR chi2(1)      =      102.19
                                                    Prob > chi2     =       0.0000
Log likelihood = -144.83762                       Pseudo R2      =       0.2608
```

	art	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
agec		1.094428	.0119299	8.28	0.000	1.071293 1.118061
_cons		.1176387	.0238726	-10.55	0.000	.0790341 .1751

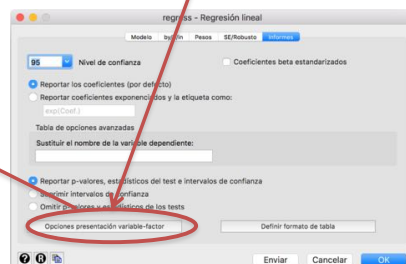
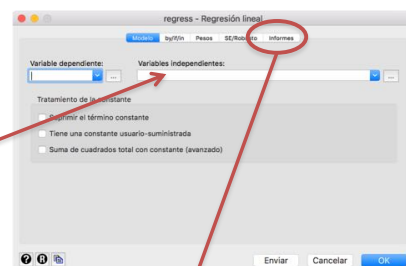
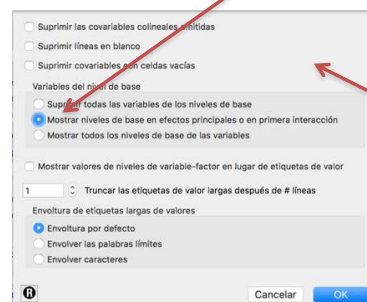
Note: \_cons estimates baseline odds.

## Modelos de regresión

- Las variables categóricas se deben ajustar utilizando el operador **i.** antes del nombre de la variable

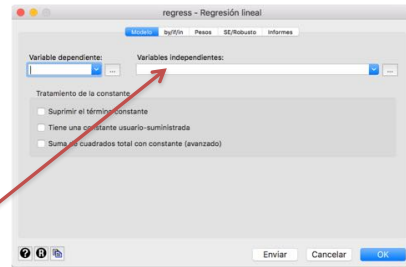
> Estadísticas > Modelos lineales y afines  
> Regresión lineal

```
. regress vary i.varx, base
```



# Modelos de regresión

- Por defecto, la categoría de referencia es la codificada con el valor más bajo
- Sin embargo, se puede cambiar utilizando el operador **ib#**. antes del nombre de la variable.



> Estadísticas > Modelos lineales y afines  
 > Regresión lineal

. regress vary **ib#.varx,** base

```
. regress bone i.sex, base
```

Source	SS	df	MS	Number of obs	=	400
Model	.921376099	1	.921376099	F(1, 398)	=	150.70
Residual	2.43342562	398	.006114135	Prob > F	=	0.0000
				R-squared	=	0.2746
				Adj R-squared	=	0.2728
Total	3.35480172	399	.008408024	Root MSE	=	.07819

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sex					
male	0	(base)			
female	-.0969431	.0078971	-12.28	0.000	-.1124683 - .0814179
_cons	.5937326	.0059622	99.58	0.000	.5820113 .6054538

```
. regress bone ib1.sex, base
```

Source	SS	df	MS	Number of obs	=	400
Model	.921376099	1	.921376099	F(1, 398)	=	150.70
Residual	2.43342562	398	.006114135	Prob > F	=	0.0000
				R-squared	=	0.2746
				Adj R-squared	=	0.2728
Total	3.35480172	399	.008408024	Root MSE	=	.07819

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sex					
male	.0969431	.0078971	12.28	0.000	.0814179 .1124683
female	0	(base)			
_cons	.4967895	.0051785	95.93	0.000	.4866089 .50697



```
. tabodds art sex
```

sex	cases	controls	odds	[95% Conf. Interval]	
male	16	156	0.10256	0.06131	0.17157
female	61	167	0.36527	0.27244	0.48973

```
. tabodds art sex , or
```

sex	Odds Ratio	chi2	P>chi2	[95% Conf. Interval]	
male	1.000000	.	.	.	.
female	3.561377	19.16	0.0000	1.939519	6.539460

```
. logistic art i.sex, base
```

```
Logistic regression                               Number of obs   =       400
                                                    LR chi2(1)      =       20.56
                                                    Prob > chi2     =       0.0000
Log likelihood = -185.65262                       Pseudo R2      =       0.0525
```

art	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
sex						
male	1 (base)					
female	3.561377	1.076053	4.20	0.000	1.969843	6.438793
_cons	.1025641	.0269239	-8.68	0.000	.0613124	.1715704

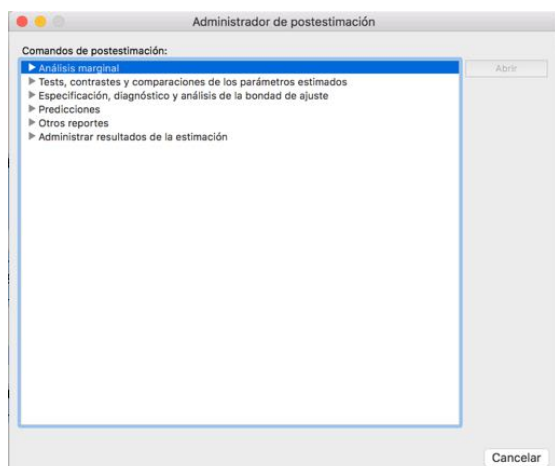
Note: \_cons estimates baseline odds.

## Post-estimación

- Después del ajuste de cualquier modelo de regresión, hay disponible un conjunto de comandos útiles de post-estimación

> Estadísticas > Post-estimación

- Estos comandos permiten,
  - Guardar nuevas variables como residuos, predicciones, etc.
  - Medidas de bondad de ajuste
  - Combinación de parámetros
  - Guardar estimaciones
  - Calcular efectos marginales



# Regresión múltiple

- **Stata** estima, de manera predeterminada, un modelo aditivo al definir múltiples variables independientes

```
. regress vary varx1 i.varx2
```
  - Para estimar un modelo multiplicativo hay que definir la interacción entre las variables independientes, utilizando los operadores #

```
. regress vary var1 i.var2 c.varx1#i.varx2
```
  - o alternativamente ##

```
. regress vary c.varx1##i.varx2
```
- Tener en cuenta que para definir el término de interacción, las variables continuas deben identificarse con el operador **c**.

19

```
. regress bone height i.sex c.height#i.sex , base
```

Source	SS	df	MS	Number of obs	=	400
Model	1.10039152	3	.366797173	F(3, 396)	=	64.43
Residual	2.2544102	396	.005692955	Prob > F	=	0.0000
				R-squared	=	0.3280
				Adj R-squared	=	0.3229
Total	3.35480172	399	.008408024	Root MSE	=	.07545

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
height	.0016565	.0008249	2.01	0.045	.0000347 .0032783
sex					
male	0	(base)			
female	-.4437274	.1861352	-2.38	0.018	-.8096641 -.0777907
sex#c.height					
female	.0023073	.0011197	2.06	0.040	.000106 .0045085
_cons	.3093513	.1417358	2.18	0.030	.0307025 .588

```
. regress bone c.height##i.sex , base
```

Source	SS	df	MS	Number of obs	=	400
Model	1.10039152	3	.366797173	F(3, 396)	=	64.43
Residual	2.2544102	396	.005692955	Prob > F	=	0.0000
Total	3.35480172	399	.008408024	R-squared	=	0.3280
				Adj R-squared	=	0.3229
				Root MSE	=	.07545

bone	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
height	.0016565	.0008249	2.01	0.045	.0000347 .0032783
sex					
male	0 (base)				
female	-.4437274	.1861352	-2.38	0.018	-.8096641 -.0777907
sex#c.height					
female	.0023073	.0011197	2.06	0.040	.000106 .0045085
_cons	.3093513	.1417358	2.18	0.030	.0307025 .588

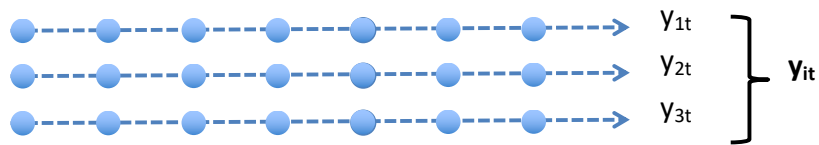
## 6. Análisis de supervivencia

## Diseños con tiempo

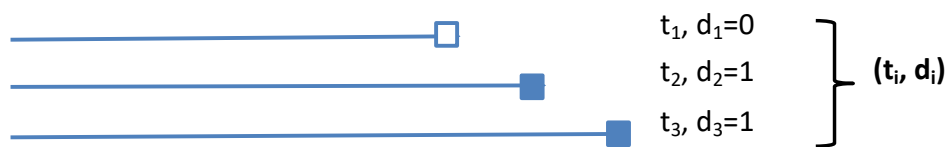
- Series temporales



- Medidas repetidas



- Análisis de supervivencia



23

## Tipos de fechas y formatos

- Stata** tiene disponibles la mayoría de los tipos de fechas cada uno tiene su propio formato
- Una fecha es una variable numérica que cuenta unidades de tiempo desde la fecha de referencia **1st Jan 1960**

Date type	Format	Examples	Time since 01jan1960
date	%td	01jan2010	18,282 days
weekly date	%tw	2010w1	2,601 weeks
monthly date	%tm	2010m1	600 months
quarterly date	%tq	2010q1	200 quarters
half-yearly date	%th	2010h1	100 half-years
yearly date	%ty	2010	since year 0

24

## Tipos de fechas y formatos

- Se pueden realizar operaciones matemáticas con variables de fecha recogidas en las mismas unidades de tiempo
- El resultado estará en la misma unidad de tiempo

Date type	Examples
-----	
<code>date2 - date1</code>	= days between date1 and date2
<code>week2 - week1</code>	= weeks between week1 and week2
<code>month2 - month1</code>	= months between month1 and month2
<code>half2 - half1</code>	= half-years between half1 and half2
<code>year2 - year1</code>	= years between year1 and year2
-----	

25

## Generar fechas

- Las variables de fecha pueden generarse a partir de otras variables que contienen diferentes unidades de tiempo
- Una vez generada hay que definir el formato apropiado

Date type	Function	+ Format
-----		
<code>date</code>	<code>td = mdy(M, D, Y)</code>	<code>%td</code>
<code>weekly date</code>	<code>tw = yw(Y, W)</code>	<code>%tw</code>
<code>monthly date</code>	<code>tm = ym(Y, M)</code>	<code>%tm</code>
<code>quarterly date</code>	<code>tq = yq(Y, Q)</code>	<code>%tq</code>
<code>half-yearly date</code>	<code>th = yh(Y, H)</code>	<code>%th</code>
<code>yearly date</code>	<code>ty = y(Y)</code>	<code>%ty</code>
-----		

26

## Generar fechas

- Las fechas también se generan a partir de variables de texto utilizando la función `date()` con la máscara apropiada

String date	Mask
01dec2006	"DMYh"
01-12-2006	"DMYh"
1dec2006	"DMYh"
1-12-2006	"DMYh"
01dec06	"DM20Y"
01-12-06	"DM20Y"
December 1, 2006	"MDY"
2006 Dec 01	"YMD"
2006-12-01	"YMD"
20061201	"YMD"

- Existen más funciones para codificar otras fechas con variables de texto

`weekly()`  
`monthly()`  
`quarterly()`  
`halfyearly()`  
`yearly()`

## Conversión entre fechas

- Las fechas también se pueden convertir entre unidades de tiempo, utilizando la función apropiada

From:	To:	weekly	monthly	quarterly	half-yearly	yearly
date		tw=wofd(td)	tm=mofd(td)	tq=qofd(td)	th=hofd(td)	ty=yofd(td)
weekly		td=dofw(tw)				
monthly		td=dofm(tm)				
quarterly		td=dofq(tq)				
half-yearly		td=dofh(th)				
yearly		td=dofy(ty)				

# Generar componentes de fecha

- A partir de una variable con formato de fecha ya existente en la base de datos, se pueden extraer otros componentes de fecha para cualquier unidad de tiempo

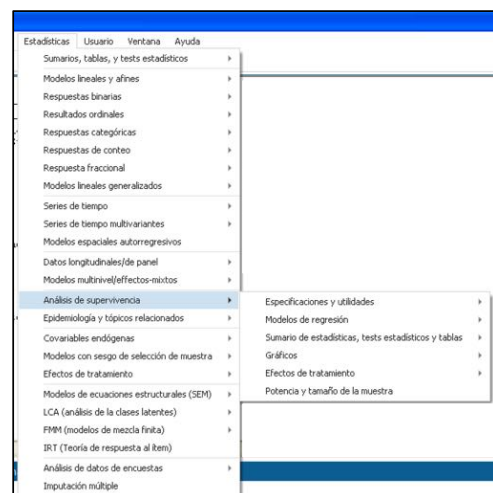
Date component	Function
calendar year	<code>yy = year (td)</code>
calendar month	<code>mm = month (td)</code>
calendar day	<code>dd = day (td)</code>
day of week (0=Sunday)	<code>dw = dow (td)</code>
week within year	<code>wy = week (td)</code>
quarter within year	<code>qy = quarter (td)</code>
half within year	<code>hy = halfyear (td)</code>

29

# Análisis de supervivencia

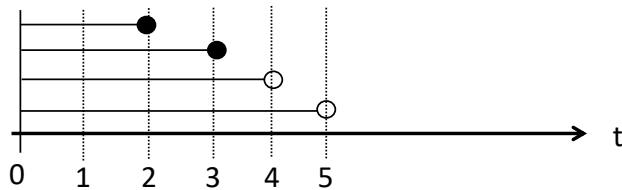
- Conjunto muy completo de comandos para el análisis de supervivencia
  - Administración de datos
  - Estadística descriptiva
  - Gráficos
  - Modelos de regresión

> *Estadísticas* > *Análisis de Supervivencia*
- Todos los comandos empiezan por **st**

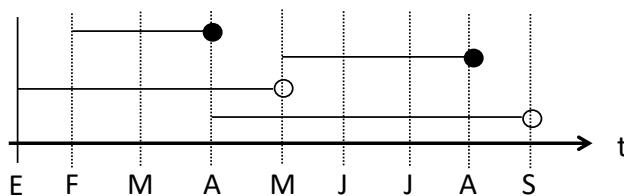


# Escalas temporales

- Cohorte fija – tiempo de seguimiento



- Cohorte dinámica – tiempo de calendario



# Definir estructura temporal

- La estructura de datos para el análisis de supervivencia debe definirse antes de utilizar cualquier comando `st`

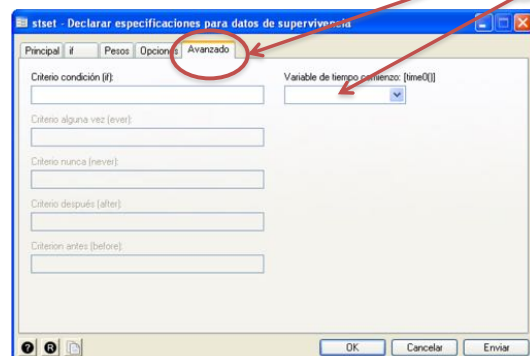
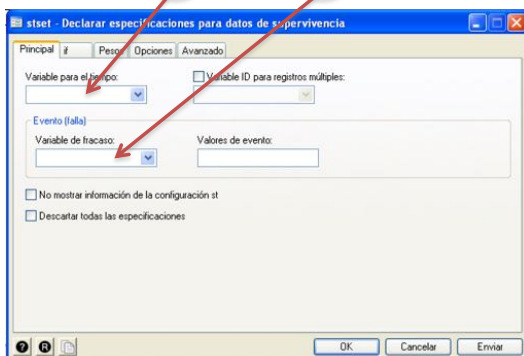
> Estadísticas > Análisis de Supervivencia > Especificaciones y utilidades > Declarar estructura de datos de supervivencia

- Tiempo de seguimiento

- Tiempo de calendario

• `stset t, fail(d)`

• `stset fs, fail(d) enter(fe)`





```

. use stanford, clear

. generate time = dox-doe
. stset time, fail(died)

    failure event: died != 0 & died < .
obs. time interval: (0, time]
exit on or before: failure
-----
    103 total observations
     0 exclusions
-----
    103 observations remaining, representing
     75 failures in single-record/single-failure data
 31,852 total analysis time at risk and under observation
                                     at risk from t =          0
                                     earliest observed entry t =      0
                                     last observed exit t =      1,799

. stset dox , fail(died) time0(doe)

    failure event: died != 0 & died < .
obs. time interval: (doe, dox]
exit on or before: failure
-----
    103 total observations
     0 exclusions
-----
    103 observations remaining, representing
     75 failures in single-record/single-failure data
 31,852 total analysis time at risk and under observation
                                     at risk from t =          0
                                     earliest observed entry t =    2,812
                                     last observed exit t =    5,204

```

## Definir estructura temporal

- Una vez definida estructura temporal de los datos para el análisis de supervivencia, **Stata** genera las siguientes variables del Sistema

```

_d      event or censored
_t0     entry time
_t      exit time
_st     record included for analysis

```

```
. describe _*
```

variable name	storage type	display format	value label	variable label
_st	byte	%8.0g		
_d	byte	%8.0g		
_t	int	%10.0g		
_t0	byte	%10.0g		

```
. stset time, fail(died)
```

```
. list id doe dox died _t0 _t _d _st in 1
```

	id	doe	dox	died	_t0	_t	_d	_st
1.	1	15 Nov 67	03 Jan 68	Died	0	49	1	1

```
. list id doe dox died _t0 _t _d _st in 1
```

```
. stset dox , fail(died) time0(doe)
```

	id	doe	dox	died	_t0	_t	_d	_st
1.	1	15 Nov 67	03 Jan 68	Died	2875	2924	1	1

## Más Escalas temporales

- A partir de fechas de calendario

The diagram illustrates the specification of survival data using calendar dates. A horizontal timeline is shown with months labeled E, F, M, A, M, J, J, A, S. Three red circles mark specific time points on the timeline, with red arrows pointing to two software dialog boxes. The top dialog box, titled 'stset - Declarar especificaciones para datos de supervivencia', shows the 'Principal' tab with 'Variable para el tiempo' set to 'dox' and 'Evento (falla)' set to 'died'. The bottom dialog box shows the 'Pesos' tab with 'Variable origen' set to 'doe' and 'Expresión tiempo origen' set to 'time0(doe)'. The 'Expresión tiempo origen' field in the bottom dialog is highlighted with a red arrow pointing to the 'M' month on the timeline.

```
. stset dox, fail(died) enter(doe) origin(dob) scale(365.25)
```

```

failure event:   died != 0 & died < .
obs. time interval: (origin, dox]
enter on or after: time doe
exit on or before: failure
t for analysis:   (time-origin)/365.25
origin:           time dob

```

```
-----
103 total observations
  0 exclusions
-----
```

```

103 observations remaining, representing
 75 failures in single-record/single-failure data
87.206 total analysis time at risk and under observation
              at risk from t =                0
              earliest observed entry t =    8.785763
              last observed exit t =       64.61602

```

```
. list id doe dox died _t0 _t _d _st in 1
```

```

-----+-----
| id          doe          dox   died   _t0          _t          _d          _st |
-----+-----
1. | 1    15 Nov 67    03 Jan 68   Died   30.844627   30.978782   1          1 |
-----+-----

```

## Tasas y regresión de Poisson

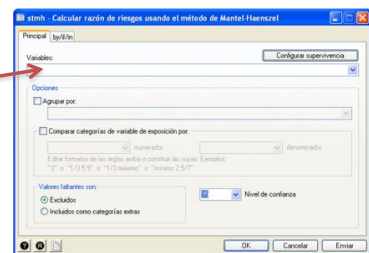
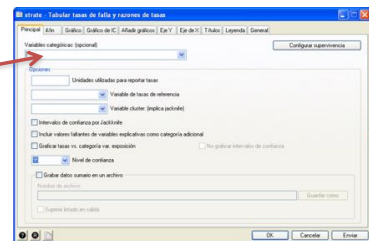
- Una vez definida estructura temporal de los datos se pueden estimar y comparar tasas de incidencia

> Estadísticas > Análisis de Supervivencia > Sumario de estadísticas, tests y tablas > Tabular tasas de falla y razones de tasas

```
. strate varx
```

> Estadísticas > Análisis de Supervivencia > Sumario de estadísticas, tests y tablas > Tabular razones de riesgos

```
. stmh varx
```



```
. stset time, fail(died) scale(365.25)
```

```
. strate
```

Estimated rates and lower/upper bounds of 95% confidence intervals  
(103 records included in the analysis)

D	Y	Rate	Lower	Upper
75	87.2060	0.86003	0.68585	1.07846

```
. strate surg
```

Estimated rates and lower/upper bounds of 95% confidence intervals  
(103 records included in the analysis)

surg	D	Y	Rate	Lower	Upper
No	66	62.1930	1.06121	0.83373	1.35076
Yes	9	25.0130	0.35981	0.18722	0.69153

```
. stmh surg
```

Maximum likelihood estimate of the rate ratio  
comparing surg==1 vs. surg==0

RR estimate, and lower and upper 95% confidence limits

RR	chi2	P>chi2	[95% Conf. Interval]
0.339	10.20	0.0014	0.169 0.680

## Tasas y regresión de Poisson

- Hay una amplia variedad de modelos paramétricos para el análisis de datos de supervivencia

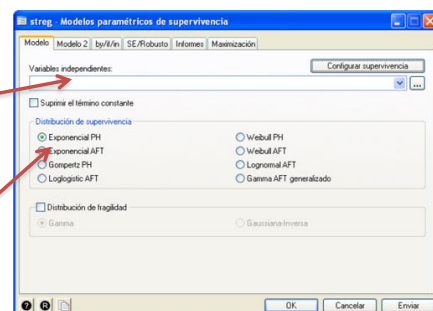
> Estadísticas > Análisis de Supervivencia >  
Modelos de regresión > Modelos de supervivencia paramétricos

```
. streg varx
```

- El modelo exponencial se corresponde a la regresión de Poisson

```
. streg varx , d(exponential)
```

- Los parámetros se muestran, por defecto, en escala multiplicativa para su directa interpretación, en forma de razones de tasas



```
. streg, d(exponential)
```

```
...
```

```
Exponential regression -- log relative-hazard form
```

```
No. of subjects =          103          Number of obs =          103
No. of failures =           75
Time at risk   =  87.20602327
Log likelihood = -234.83377          LR chi2(0) =          -0.00
                                      Prob > chi2 =           .
```

```
-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      _cons |   .8600323   .099308   -1.31   0.192   .6858458   1.078458
-----+-----
```

```
. generate tyear = time/365.25
```

```
. poisson died, exp(tyear) irr
```

```
Poisson regression          Number of obs =          103
                              LR chi2(0) =           0.00
                              Prob > chi2 =           .
Log likelihood = -234.83377   Pseudo R2 =          0.0000
```

```
-----+-----
      died |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      _cons |   .8600323   .099308   -1.31   0.192   .6858458   1.078458
ln(tyears) |           1 (exposure)
-----+-----
```

```
. streg i.surg, d(exponential) base
```

```
...
```

```
Exponential regression -- log relative-hazard form
```

```
No. of subjects =          103          Number of obs =          103
No. of failures =           75
Time at risk   =  87.20602327
Log likelihood = -228.80323          LR chi2(1) =          12.06
                                      Prob > chi2 =          0.0005
```

```
-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      surg |
      No  |           1 (base)
      Yes |   .3390583   .1204791   -3.04   0.002   .1689713   .6803551
      _cons |  1.061212   .1306262    0.48   0.629   .8337323   1.350759
-----+-----
```

```
. display 1.061212*.3390583
```

```
.35981274
```

# Supervivencia y regresión de Cox

- **Stata** permite **listar, graficar y comparar** curvas de supervivencia y riesgo acumulado, estimadas por el método de Kaplan-Meier
  - > Estadísticas > Análisis de Supervivencia > Sumario de estadísticas, tests y tablas > **Listar función de supervivencia**

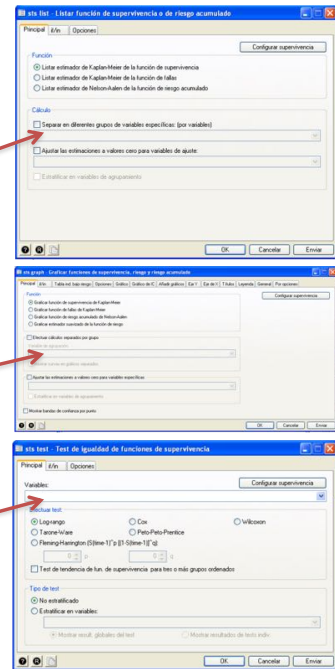
`. sts list varx`

- > Estadísticas > Análisis de Supervivencia > Gráficos > **Función de supervivencia de Kaplan-Meier**

`. sts graph varx`

- > Estadísticas > Análisis de Supervivencia > Sumario de estadísticas, tests y tablas > **test de igualdad de funciones de supervivencia**

`. sts test varx`



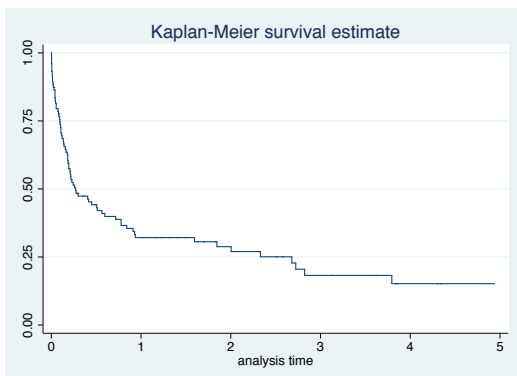
```
. sts list
...

. sts list, at(0(1)5)

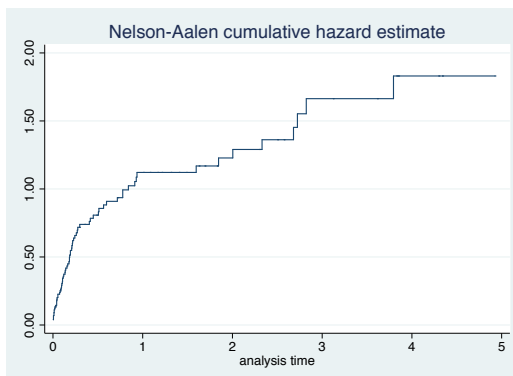
failure_d: died
analysis time_t: time/365.25
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]
0	0	0	1.0000	.	.
1	29	67	0.3212	0.0477	0.2305 0.4153
2	17	2	0.2879	0.0483	0.1976 0.3844
3	9	5	0.1823	0.0489	0.0985 0.2865
4	4	1	0.1519	0.0493	0.0713 0.2606
5	1	0	.	.	.

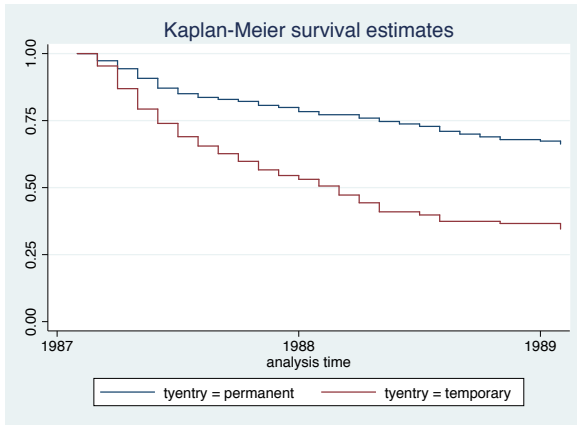
```
. sts graph
```



```
. sts graph, na
```



```
. sts graph, by(prev_surg)
```



```
. sts test prev_surg
```

```
...
```

prev_surg	Events observed	Events expected
No	66	58.56
Yes	9	16.44
Total	75	75.00

chi2(1) = 4.47  
Pr>chi2 = 0.0345

## De regresión de Poisson a Cox

- Regresión de Poisson

- $\log(\lambda) = \alpha + \beta X$

- Asume tasa ( $\lambda$ ) constante durante el período de seguimiento, aunque podemos modelizar cambios en el tiempo indicándolo explícitamente en el modelo

- Regresión de Cox

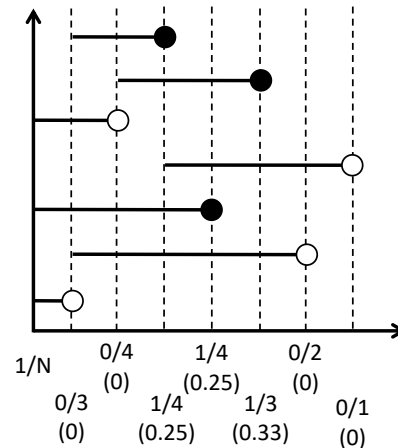
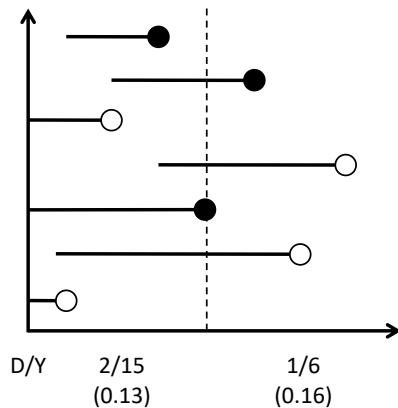
- Permite que tasas cambien rápidamente en el tiempo ( $\lambda_t$ ) y estima la razón de tasas (*hazard ratio*) controlando por el tiempo

- $\log(\lambda_t) = \log(\lambda_{0t}) + \beta X \Rightarrow \log(\lambda_t / \lambda_{0t}) = \beta X$

- Proporcionalidad de riesgos ( $\lambda_t / \lambda_{0t}$  no depende del tiempo)

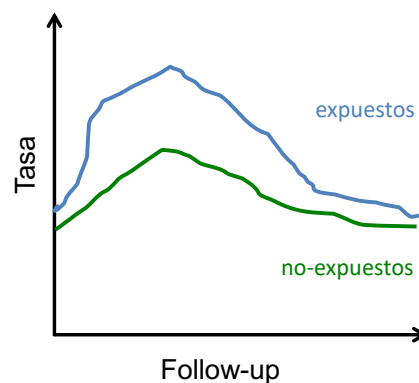
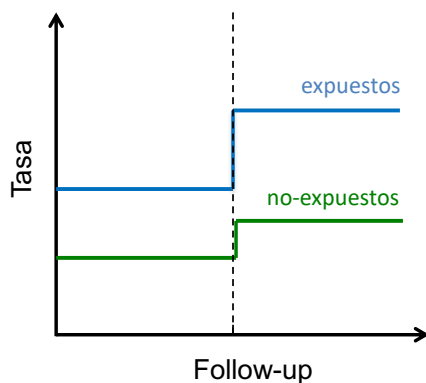
## De regresión de Poisson a Cox

- En regresión de Poisson tasa constante =  $D/Y$  en cada período de tiempo
- En regresión de Cox se definen instantes de tiempo con un único evento, hazard rate =  $1/N$



## De regresión de Poisson a Cox

- Regresión de Poisson asume tasa constante y permite indicar que la tasa cambian en el tiempo
- En regresión de Cox la tasa basal cambia en el tiempo y asume riesgos (hazards) proporcionales





# Supervivencia y regresión de Cox

- La regresión de Cox está incluida entre la variedad de modelos de regresión para análisis de datos de supervivencia

> Estadísticas > Análisis de Supervivencia > Modelos de regresión > Modelo de riesgos proporcionales de Cox

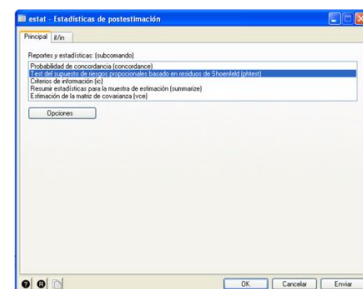
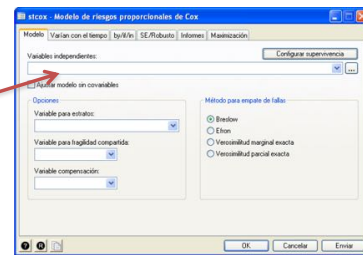
. **stcox varx**

- Los parámetros se muestran, por defecto, en escala multiplicativa para su directa interpretación, en forma de *Hazard Ratio*

- Una vez estimado el modelo de Cox se debe comprobar la asunción de proporcionalidad de riesgos

> Estadísticas > Análisis de Supervivencia > Modelos de regresión > Test para el supuesto de riesgos proporcionales

. **estat phtest**



```
. stcox i.surg, base
```

```
...
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
surg						
No	1	(base)				
Yes	.4766721	.1711623	-2.06	0.039	.2358155	.9635342

```
. estat phtest
```

Test of proportional-hazards assumption

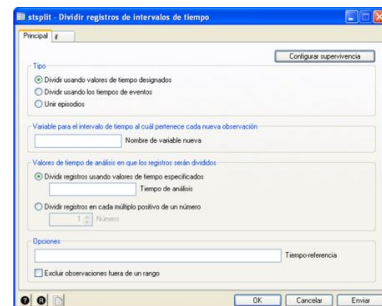
Time: Time

	chi2	df	Prob>chi2
global test	1.73	1	0.1881

## ¿Qué hacer?

- *Mejorar* la especificación del modelo de regresión incluyendo otras variables confusoras
- Estratificar regresión de Cox o incluir interacción tiempo  $\times$  exposición en regresión de Poisson
- Dividir el tiempo de seguimiento en pacientes con información que cambia en el tiempo (expansión Lexis)

> Estadísticas > Análisis de Supervivencia > Especificaciones y utilidades > Partir períodos de tiempo en sub-períodos



## VARIABLES TIEMPO-DEPENDIENTES

- Cambio de no expuesto a expuesto durante el tiempo de seguimiento (o viceversa)

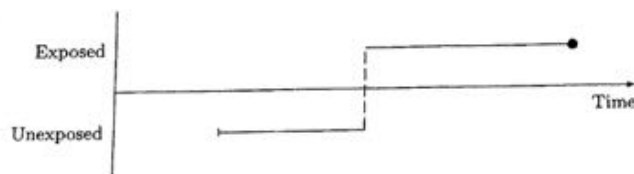


Fig. 31.1. Changing exposure group.

- *Dividir* tiempo de seguimiento para cada paciente en dos partes: períodos de no exposición y de exposición
- Incluyendo estas partes como distintos pacientes en el modelo de regresión de Cox

## Variables tiempo-dependientes

- Escalas de tiempo 'complementarias' como variables explicativas

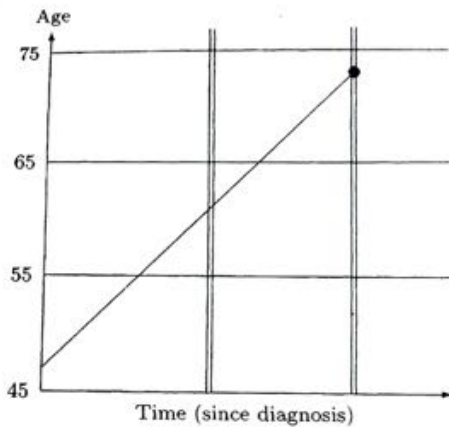


Fig. 31.2. Follow-up by age and time.

- En seguimientos breves considerar edad diagnóstico
- En seguimientos mas largos considerar edad que por si misma cambia en el tiempo
- Dividir edad en grupos de 5- o 10- años como variable categórica

## Resumen

- Principales comandos para modelos de regresion
  - . `pwcorr`
  - . `graph twoway () () ...`
  - . `regress, logistic, poisson`
- Principales comandos para análisis de supervivencia
  - . `stset`
  - . `strate, stmh, streg`
  - . `sts [list | graph | test], stcox`