Regresión de Series Temporales

Dr. Marcelo Risk

Data Mining de Series Temporales, Maestría en Explotación de Datos y Descubrimiento de Conocimientos, FCEyN UBA

Mayo 2016

Métodos estadísticos más utilizados de acuerdo a la escala de las variables

Escala variable dependiente	**Escala variables independientes**	**Método estadístico**
Intervalar	Intervalar	Regresión, múltiple en el caso de más de una variable independiente
Intervalar	Nominal e intervalar	Análisis de la covarianza (ANCOVA)
Intervalar	Nominal u ordinal	Análisis de la varianza simple (ANOVA), múltiple (MANOVA)
Dicotómica	Nominal e intervalar	Regresión logística, simple o múltiple

Análisis de regresión

- El análisis de regresión fue introducido por Francis Galton (1822-1911), científico británico de la época victoriana, el cual contribuyó al conocimento de la antropometría, la psicología diferencial, la geografía y la estadística en tre otras; además fue primo de Charles Darwin.
- Uno de sus trabajos más influentes es *Regression Towards Mediocrity in Hereditary Stature*, publicado en 1886 en el Journal of the Anthropological Institute; en dicho trabajo Francis Galton analizó graficamente la relación entre la altura de padres e hijos, concluyendo que la altura media de hijos nacidos de padres de una dada altura tienden a valores de la media de la población, luego Galton explicó esto diciendo que fue una regresión a los valores medios de la población.

paper de Francis Galton

ANTHROPOLOGICAL MISCELLANEA.

Regression *towards* Mediocrity *in* Hereditary Stature.

By Francis Galton, F.R.S., &c.

[With Plates IX and X.]

This memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here the portion of it which bears upon regression, together with some amplification where brevity had rendered it obscure, and I have added copies of the diagrams suspended at the meeting, without which the letterpress is necessarily difficult to follow. My object is to place beyond doubt the existence of a simple and far-reaching law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees. I once before ventured to draw attention to this law on far more slender evidence than I now possess.

Análisis de regresión

- El análisis de regresión se puede utilizar para describir la relación, su extensión, dirección e intensidad, entre una o varias variables independientes con escala intervalar y una variable dependiente tambien intervalar.
- Cabe destacar que no debe utilizarse el análisis de regresión como prueba de causalidad (causa - efecto), en realidad no hay métodos estadísticos para probar causalidad.

Análisis de regresión

La correcta aplicación del análisis de regresión asume que se cumplen las siguientes condiciones:

a) **Existencia**: para cada valor fijo de X, existe un valor Y aleatorio con una distribución de probabilidad con valores finitos de media y varianza, esta condición debe cumplirse siempre.

b) **Independencia**: los valores Y son estadisticamente independientes uno de otro, esta condición puede no cumplirse en el caso de varias observaciones (valores de Y) sobre un mismo valor de X, en ese caso debe tenerse en cuenta dicha condición.

c) **Linearidad**: el valor medio de Y es una función lineal de X.

d) **Distribución normal**: por cada valor fijo de X, Y tiene una distribución normal.

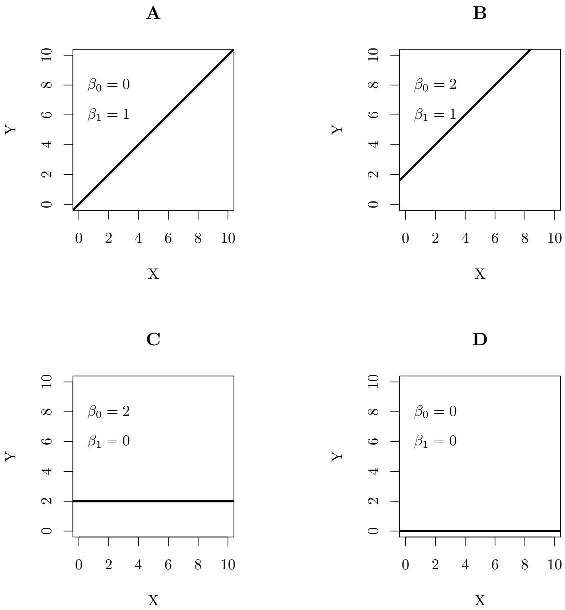
Análisis de regresión

En el análisis de regresión moderno, en su forma más simple es decir lineal y con una sola variable independiente, la variable independiente X re relaciona con la variable dependiente Y de acuerdo con la siguiente ecuación 1:

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$

donde ϵ son los residuos, es decir la diferencia entre la estimación y los valores reales para cada punto XY. Los coeficientes de la ecuación 1 son: β_0 denominado intersección (cuando $X = 0$) y β_1 pendiente.

Análisis de regresión: posibles resultados



Análisis de correlación

El análisis de correlación estudia el grado ó intensidad de asociación entre dos variables.
El grado de asociación se expresa con el **coeficiente de correlación** (r), también denominado de Pearson, se define con la siguiente ecuación:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

Note que X e Y pueden intercambiar su lugar en la ecuación y el resultado es el mismo, con lo cual no se asume jerarquía ni dependencia entre las variables.

Coeficiente de correlación

El coeficiente de correlación puede tener valores dentro del rango $-1 \leq r \leq 1$, el signo indica la dirección de la asociación, por lo tanto se pueden dar tres situaciones:

- **Asociación baja**: cuando $r \approx 0$, o cerca del 0.
- **Asociación positiva**: cuando r se acerca (tiende) a 1, el incremento en una de las variables se acompaña con un incremento en la otra, idem con el decremento.
- **Asociación negativa**: cuando r se acerca (tiende) a -1, el incremento en una de las variables se acompaña con un decremento en la otra, y viceversa.

Coeficiente de correlación: prueba de hipótesis

En la estimación del valor central, es decir de la *media* donde existe una *error* de la estimación de la media [1], y con la prueba z podemos probar la H_0: $\overline{x} = 0$.
Los mismo sucede en el caso del *coeficiente de correlación* tenemos que probar la H_0: $r = 0$, una forma es calcular el estadístico t de Student, y luego la P, con la siguiente ecuación:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

para $gl = n - 2$.
La función de R *cor.test* calcula el coeficiente de correlación r, el estadístico t y la P correspondiente, y el $IC95\%$ del r.

[1] EEM: error de estimación de la media

Ejemplo de correlación de un estudio

- Un estudio sobre la actitud de compra de un producto mide la edad de dos grupos de personas, y sus ingresos.
- Las variables observadas son:
  - Edad en años (intervalar).
  - Ingresos en K$ (intervalar).
  - Actitud de Compra del producto, es un índice que surge de una encuesta (intervalar).
  - Dos grupos: C y P (nominal).
- Se estudiaron un total de 41 sujetos.

Ejemplo de correlación de un estudio [2]

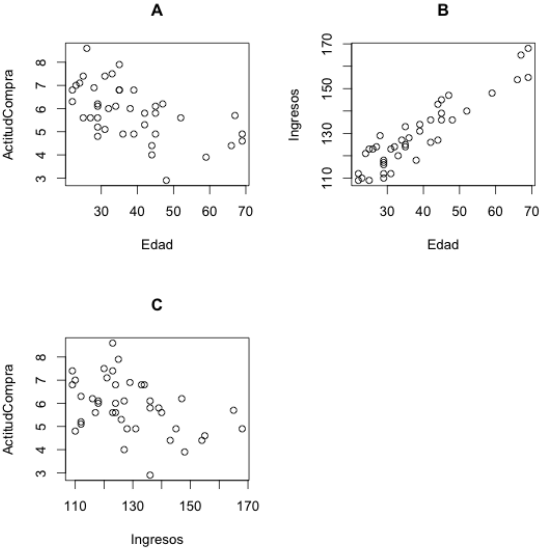
Sujeto	Edad	Ingresos	ActitudCompra	Grupo
1	22	112	6.3	A
2	22	109	6.8	B
3	23	110	7	A
4	24	121	7.1	A
5	25	123	5.6	B
6	25	109	7.4	A
7	26	123	8.6	A
8	27	124	5.6	A
9	28	129	6.9	A
10	29	112	5.2	B
11	29	118	6.1	A
12	29	117	5.6	A
13	29	110	4.8	B
14	29	116	6.2	A
...	...	...	...	...
40	69	168	4.9	A
41	69	155	4.6	A

[2] archivo *EstudioEdadIngresos.csv*

Ejemplo de correlación de un estudio

```r
dir='/ .... /' # directorio del csv

dato=read.csv(paste( dir , 'EstudioEdadIngresos.csv' ,sep=''))
print (names(dato))
attach(dato)

op=par(mfrow=c(2,2))
plot(Edad,ActitudCompra,main='A')
plot(Edad,Ingresos,main='B')
plot( Ingresos ,ActitudCompra,main='C')

print (cor.test(Edad,Ingresos))
print (cor.test(Edad,ActitudCompra))
print (cor.test( Ingresos ,ActitudCompra))
```

Actitud de compra vs Edad (A), Ingresos vs Edad (B), y Actitud de Compra vs Ingresos (C)



Ejemplo de correlación de un estudio: *cor.test* para Edad vs Ingresos

```
Pearson's product-moment correlation

data:  Edad and Ingresos
t = 14.8202, df = 39, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8567785 0.9576694
sample estimates:
      cor
0.9215256
```

Análisis de regresión del estudio

Entonces si aplicamos el análisis de regresión al estudio de Actitud de compra, obtenemos la siguiente ecuación:

$$ActitudCompra \approx \beta_0 + \beta_1 Edad + \epsilon$$

En este análisis se asume que la *Actitud de Compra* depende de la *Edad*, se relacionan de esa forma. No podría ser al revés, es decir que la *Edad* dependiera de la *Actitud de Compra*, lo podemos demostrar por el absurdo: si por ejemplo una persona cambia su *Actitud de Compra*, le cambiaría su *Edad* ?

Análisis de regresión del estudio: programa en R

```
dir = '/ .... /' # directorio del csv

dato=read.csv(paste(dir,'EstudioEdadIngresos.csv',sep=''))
print(names(dato))
attach(dato)

plot(Edad,ActitudCompra)
reg=lm(ActitudCompra~Edad)
abline(reg)

print(summary(reg))
```

Análisis de regresión del estudio

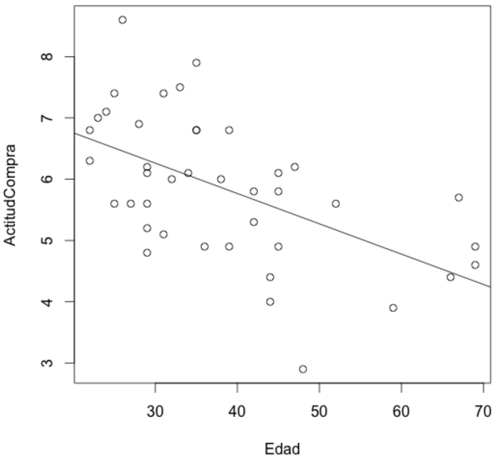


Figura: Actitud de compra vs Edad, con la correspondiente recta de regresión.

Análisis de regresión del ejemplo

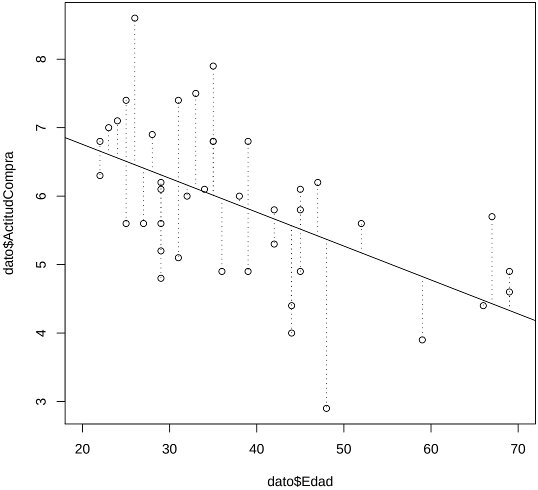
	Estimate	Std. Error	t value	P
(Intercept)	7.74601	0.48590	15.942	< 2e-16 ***
Edad	-0.04951	0.01203	-4.116	0.000193 ***

La estimación de la pendiente es estadísticamente significativa ($P = 0{,}0002$) con lo cual podemos concluir que existe una asociación entre la Edad y la Actitud de Compra, el signo negativo de la estimación de la pendiente nos indica que a medida que se incrementa la *Edad* la *Actitud de Compra* se decrementa, a un ritmo de $-0{,}04951$ por año, el coeficiente de intersección es estadísticamente significativo (distinto de cero), con una estimación de 7,74, dicho valor toma la Actitud de Compra cuando la $Edad = 0$.

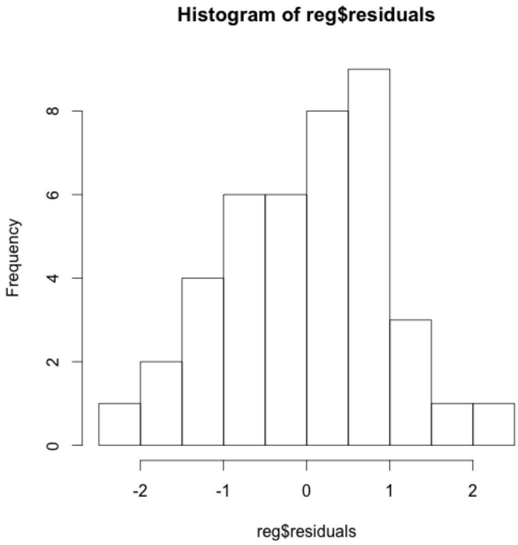
Análisis de regresión Actitud de compra vs Edad

```
dir='/ .... /' # directorio del csv
dato=read.csv(paste(dir,'EstudioEdadIngresos.csv',sep=''))
print(names(dato))
attach(dato)
reg=lm(ActitudCompra~Edad)
plot(Edad,ActitudCompra,xlim=c(20,70))
abline(reg)
interseccion = reg$coefficients[1]
pendiente = reg$coefficients[2]
n=length(Edad)
for(i in 1:n){
  xx <- Edad[i]
  y1 <- interseccion+pendiente*xx
  y2 <- y1+reg$residuals[i]
  lines(c(xx,xx),c(y1,y2),lty='dotted')
}
hist(reg$residuals)
```

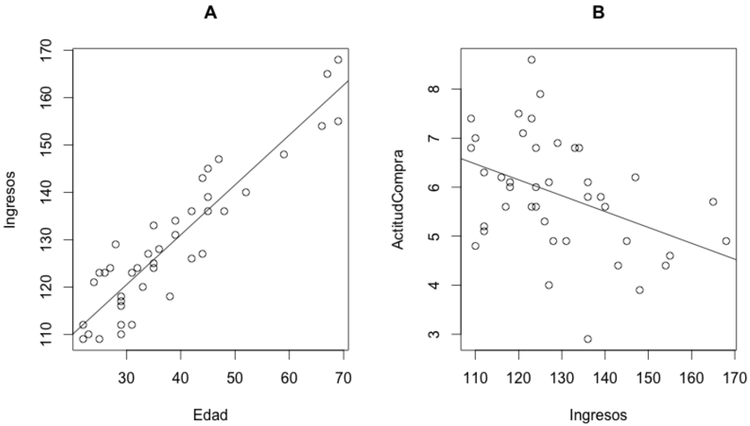
Actitud de compra vs Edad, mostrando en linea de puntos los residuos a la recta de regresión



Actitud de compra vs Edad, histograma de los residuos



Ingresos versus Edad (A), y Actitud de Compra versus Ingresos (B), ambos con la correspondiente recta de regresión



Análisis de regresión para Ingresos versus Edad

	Estimate	Std. Error	t value	P
(Intercept)	88.8701	2.8722	30.94	<2e-16 ***
Edad	1.0537	0.0711	14.82	<2e-16 ***

- Podemos ver que la distribución de los residuos parece ser normal, luego que la estimación de los dos coeficientes es estadisticamente significativa para, finalmente los Ingresos se incrementan con la Edad a un ritmo de 1.054 K$/año, con un coeficiente de intersección de 88,9$K$$.

Análisis de regresión para Actitud de Compra versus los Ingresos

	Estimate	Std. Error	t value	P
(Intercept)	10.03862	1.49285	6.724	5.13e-08 ***
Ingresos	-0.03241	0.01148	-2.823	0.00745 **

- En este caso los dos coeficientes son estadisticamente significativos, la Actitud de Compra se decrementa con un incremento de los Ingresos con un ritmo de $-0{,}0324$ 1/K$, y un valor de intersección de 10,04.

Intervalos de confianza de la regresión

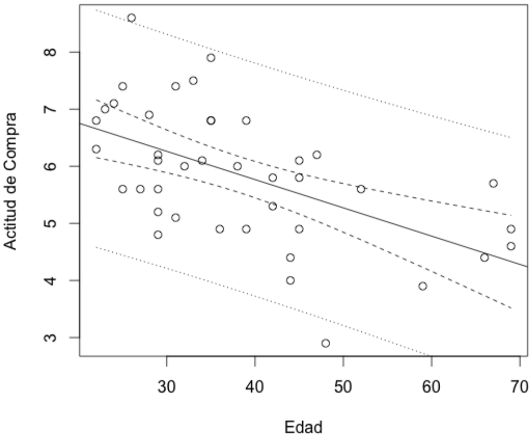
En algunas aplicaciones puede ser útil calcular el intervalo de confianza de la recta de regresión, de esta forma podemos verificar para una probabilidad dada (generalmente 95 %), hasta donde se puede extender la recta de regresión.
Los intervalos de confianza (IC) para la recta de regresión, se define como la ecuación 2:

$$Y_{IC95\%,X_0} = \overline{Y} + \beta_1(X_0 - \overline{X}) \pm t_{n-2,1-0,95/2}\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{(n-1)S_X^2}} \quad (2)$$

Por otro lado puede ser también conveniente calcular las bandas de predicción (BP), las cuales se definen como la ecuación 3:

$$Y_{BP95\%,X_0} = \overline{Y} + \beta_1(X_0 - \overline{X}) \pm t_{n-2,1-0,95/2}\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{(n-1)S_X^2}} \quad (3)$$

Actitud de Compra versus Edad, con IC95% para la recta de regresión (lineas de rayas) y las bandas de predicción del 95% (lineas de puntos)



Análisis de regresión no lineal

En algunas situaciones la relación entre dos variables no es lineal, por lo tanto la regresión lineal no describe adecuadamente dicha relación.
Si bien es posible calcular la regresión con cualquier función, una de las regresiones no lineales más comunes es la *exponencial*, cuya ecuación es la siguiente:

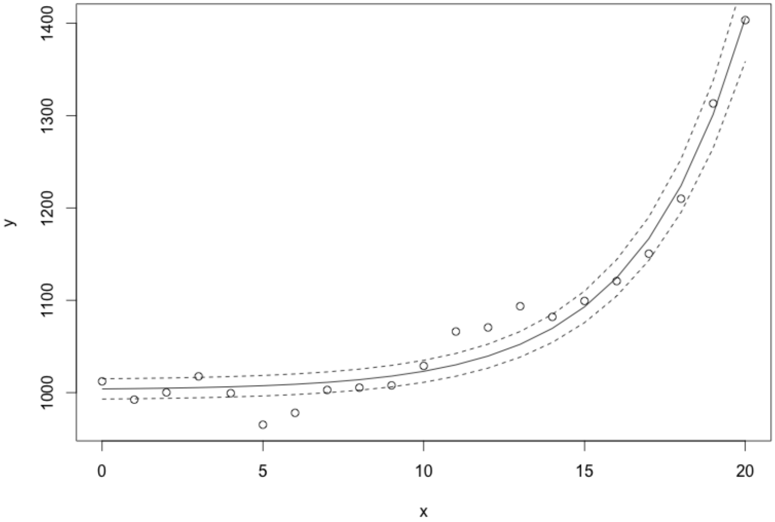
$$Y \approx \beta_0 + e^{\beta_1 X} + \epsilon$$

Los coeficientes de la ecuación de regresión son: β_0 denominado intersección al origen (cuando $X = 0$), en este caso el *offset* en Y, y β_1 el exponente; ϵ representa las diferencias entre la regresión y los datos, esta diferencia se denomina *residuos*.

Análisis de regresión no lineal

```
coef.a = 1000
coef.b = 0.3
x = 0:20
y = coef.a + exp(coef.b*x) + 20*rnorm(length(x))

plot(x,y)
lines(x,coef.a + exp(coef.b*x))

regexp =
    nls(y~a+exp(b*x),start=list(a=500,b=0.5),trace=TRUE)
print(summary(regexp))

plot(x,y)
lines(x,1003+exp(0.2999*x))
lines(x,992+exp(0.2952152*x),lty=2)
lines(x,1014+exp(0.3042097*x),lty=2)
```

Análisis de regresión no lineal



Análisis de regresión no lineal

```
> regexp = nls(y~a+exp(b*x),start=list(a=500,b=0.5),trace=TRUE)
685927799 :  500.0    0.5
93788130 :  856.9771446  0.4502208
11979928 :  943.5881583  0.4020525
1338276 :  979.527299  0.358111
106447.4 :  993.1654415  0.3237708
8132.706 :  997.1797706  0.3054622
5506.889 :  997.8147575  0.3012699
5502.443 :  997.8446155  0.3010825
5502.443 :  997.8450291  0.3010818
> print(summary(regexp))

Formula: y ~ a + exp(b * x)

Parameters:
   Estimate Std. Error t value Pr(>|t|)
a 9.978e+02 4.328e+00   230.5   <2e-16 ***
b 3.011e-01 1.716e-03   175.5   <2e-16 ***
```

Análisis de Covarianza

- En los análisis anteriores pudimos verificar la asociación entre la *Edad*, la *Actitud de Compra* y los *Ingresos*, pero no tuvimos en cuenta a que *Grupo* pertenecía cada sujeto.
- El método estadístico que permite incluir la variable Grupo, que tiene escala nominal, es el análisis de la *covarianza*.
- El coeficiente estimado de la variable en escala nominal es el más importante en este análisis, por otro lado la variable en escala intervalar se denomina *confundente*.
- El análisis de la covarianza permite estimar la diferencia entre los distintos grupos, asumiendo que hay dos rectas de regresión con la misma pendiente.

Análisis de Covarianza

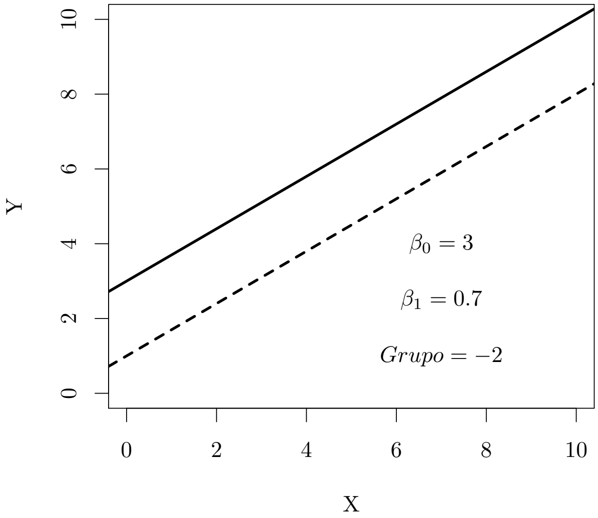
Entonces la expresión matemática del análisis de covariamza es::

$$Y \approx \beta_0 + \beta_1 X + Grupo + \epsilon$$

donde *Grupo* es la variable *confundente*, al igual que en la regresión lineal β_0 se denomina intersección al origen [3] (cuando $X = 0$) y β_1 la pendiente; ϵ representa las diferencias entre la regresión y los datos, esta diferencia se denomina *residuos*.

[3] intercept u ordenada al origen.

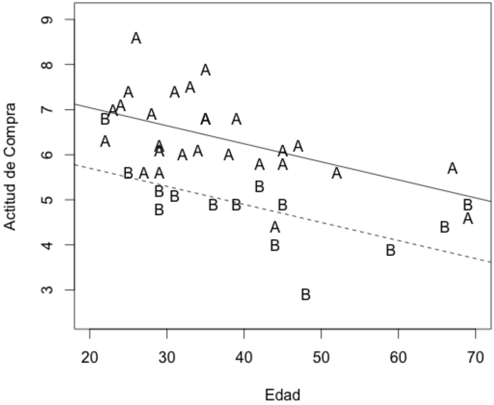
Análisis de Covarianza



Análisis de Covarianza

```
dir = '/ .... /' # directorio del csv
dato=read.csv(paste( dir , 'EstudioEdadIngresos. csv' ,sep=''))
print (names(dato))
attach(dato)
covar . ActitudCompra.Edad = lm(ActitudCompra~Edad+Grupo)
print (summary(covar.ActitudCompra.Edad))
plot (Edad[Grupo=="A"],ActitudCompra[Grupo=="A"],
      xlim=c(20,70),ylim=c(min(ActitudCompra)−0.5,max(ActitudCompra
      main="",pch='A',
      xlab="Edad",ylab="Actitud de Compra")
abline (covar . ActitudCompra.Edad$coefficients [1], covar . ActitudCompra.
points (dato$Edad[dato$Grupo=="B"],dato$ActitudCompra[dato$Grupo=
abline (covar . ActitudCompra.Edad$coefficients[1]+covar . ActitudCompra.
        covar . ActitudCompra.Edad$coefficients [2], lty ="dashed")
print ( confint (covar . ActitudCompra.Edad))
```

Actitud de Compra versus Edad con Grupo como *covariable* [4]



[4] la regresión del Grupo de A se indica con una línea llena, mientras que la del grupo de la B se indica con una línea de rayas.

Análisis de covarianza del ejemplo

	Estimate	Std. Error	t value	P
(Intercept)	7.843286	0.377348	20.785	< 2e-16 ***
Edad	-0.040070	0.009506	-4.215	0.000148 ***
GrupoB	-1.343233	0.259345	-5.179	7.57e-06 ***

- El resultado numérico del análisis de la covarianza es el siguiente, vemos que la estimación de la pendiente es -0.04 ($P < 0{,}001$), la estimación de GroupB es -1.34 ($P < 0{,}001$), con lo cual se concluye que la *Actitud de Compra* ajustada por la *Edad* de B es 1.34 menor que de A, esa es la diferencia constante entre las dos rectas de regresión a lo largo de *Edad*.

Análisis de covarianza del ejemplo: IC95% de los coeficientes

```
print(confint(covar.ActitudCompra.Edad))
```

	2.5%	97.5%
(Intercept)	7.07938484	8.60718624
Edad	-0.05931398	-0.02082553
GrupoB	-1.86824977	-0.81821596

Análisis de regresión múltiple

En los casos donde existe más de una variable independiente la regresión se convierte en *regresión múltiple*, con este método también es posible estudiar la interacción de dos variables calculando un coeficiente extra, con la ecuación:

$$ActitudCompra = \beta_0 + \beta_1 Edad + \beta_2 Ingresos + \beta_3 Edad.Ingresos + \epsilon$$

En el ejemplo se asume que la *Actitud de Compra* depende de la *Edad* y de los *Ingresos*, en forma simultánea, y de la *interacción* de ambas variables independientes.
En la *regresión múltiple* cada variable independiente *controla* a las demás, y aún en el caso que sus coeficientes no sean significativos ($P > 0{,}05$) influyen en el análisis controlando el mismo.

Análisis de regresión múltiple del ejemplo

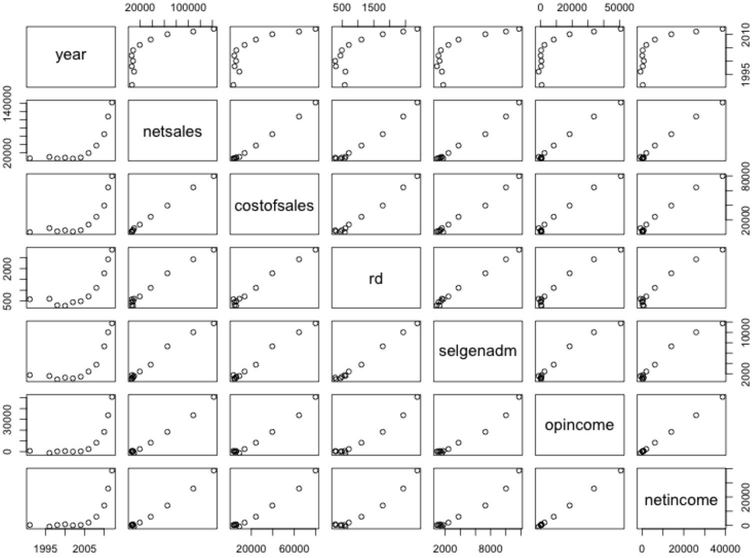
	Estimate	Std. Error	t value	P
(Intercept)	8.0639754	4.4991541	1.792	0.0813 .
Edad	-0.2293775	0.1073820	-2.136	0.0394 *
Ingresos	0.0153985	0.0380526	0.405	0.6881
Edad:Ingresos	0.0008936	0.0007230	1.236	0.2242

- El resultado nos muestra que la *Actitud de Compra* se decrementa a un ritmo de $-0{,}23$ 1/Año ($P = 0{,}039$), sin embargo los coeficientes estimados para los *Ingresos* y la interacción *Edad:Ingresos* no son significativos, podemos concluir que la *Actitud de Compra* se decrementa significativamente con el incremento de la *Edad* controlado por los *Ingresos* y la interacción de la *Edad* y los *Ingresos*.

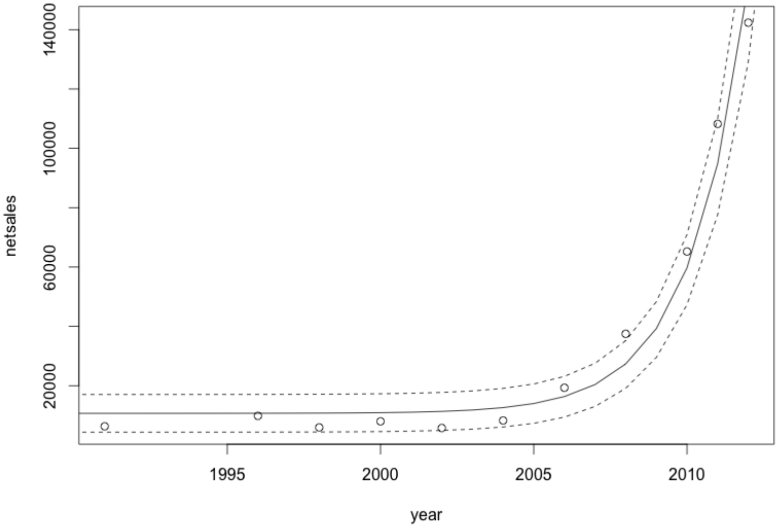
Caso de Harvard *Apple Inc. in 2012*

```
dir='/ .... /'
dato = read.csv(paste( dir , 'CasoHarvardApple2012.csv',sep=''))
print (names(dato))
attach(dato)

plot(dato)
print (cor(dato))

regexp1 =
    nls( netsales ~a+exp(b*(year-1990)),start=list(a=10000,b=0.5),trace=TRU
print (summary(regexp1))
print ( confint (regexp1))

plot( year , netsales )
year2=1990:2015
lines (year2,10690+exp(0.54*(year2-1990)))
lines (year2,4334+exp(0.5334231*(year2-1990)),lty=2)
lines (year2,17046+exp(0.5447583*(year2-1990)),lty=2)
```

Caso de Harvard *Apple Inc. in 2012*



Caso de Harvard *Apple Inc. in 2012*



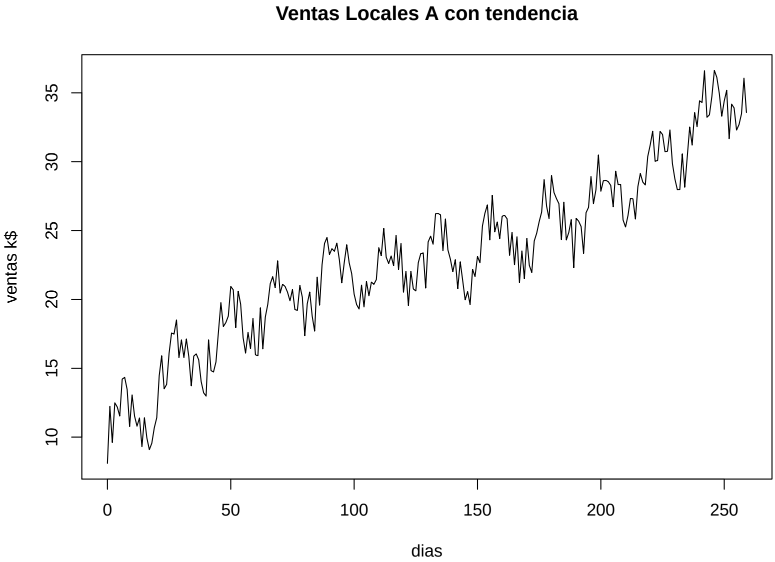
Caso de Harvard *Apple Inc. in 2012*

```
> regexp1 =
    nls( netsales~a+exp(b*(year-1990)),start=list(a=10000,b=0.5),trace=TRU
10673061128 :  1e+04 5e-01
1382120354 :  1.015829e+04 5.314709e-01
589547126 :  10654.514038     0.540219
581239586 :  1.069084e+04 5.394425e-01
581239349 :  1.068885e+04 5.394396e-01
> print(summary(regexp1))

Formula: netsales ~ a + exp(b * (year - 1990))

Parameters:
   Estimate Std. Error t value Pr(>|t|)
a 1.069e+04  2.811e+03   3.802    0.0042 **
b 5.394e-01  2.499e-03 215.886   <2e-16 ***

           2.5%        97.5%
a 4334.6553580 1.704619e+04
b    0.5334231 5.447583e-01
```

Detrending de ST con regresión

```
dir = '/Users/marcelorisk/Dropbox/MateriaDataMining/Clases/'
dos.locales = read.csv(paste(dir,'DosLocales.csv',sep=''))
print(names(dos.locales))
# [1] "X"       "t"       "localA" "localB"

m = 0.1
localAtend = dos.locales$localA + m*dos.locales$t
print(tendencia$coefficients )

plot(dos.locales$t,localAtend,type = 'l',main='Ventas Locales
    A con tendencia',xlab='dias',ylab='ventas k$')
```

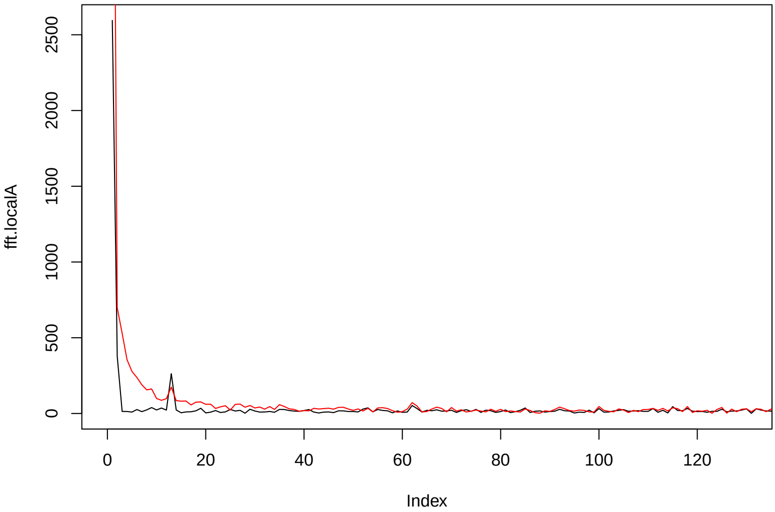
Detrending de ST con regresión



Detrending de ST con regresión

```
fft.localA = Mod(fft(dos.locales$localA))
fft.localAtend = Mod(fft(localAtend))

plot(fft.localA,type='l',xlim=c(0,length(fft.localA)/2))
lines(fft.localAtend,col='red')
```

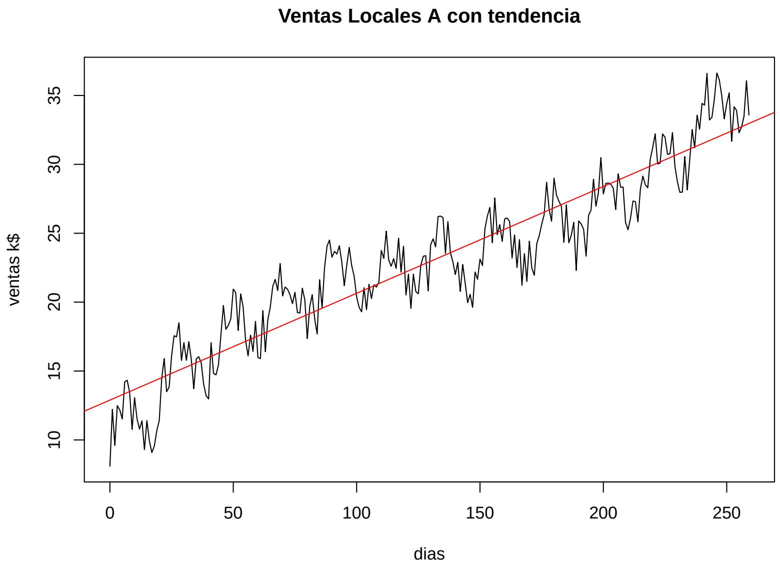
Detrending de ST con regresión



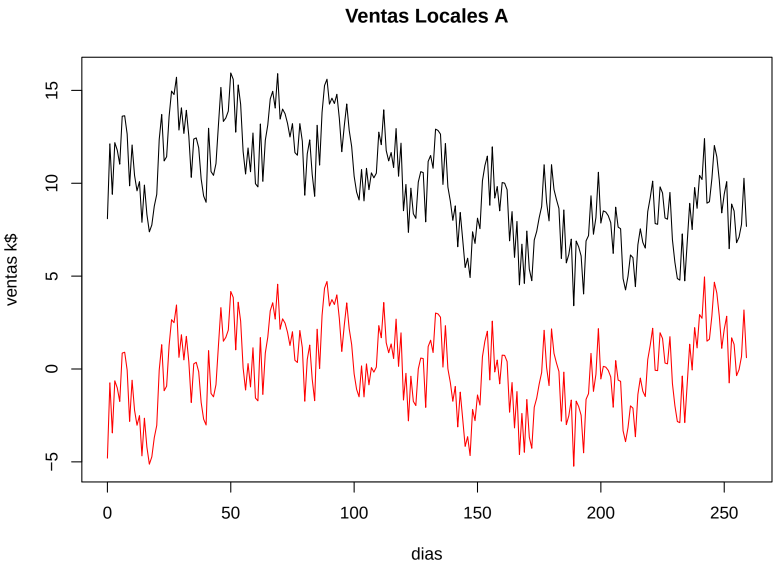
Detrending de ST con regresión

```
tendencia = lm(localAtend ~ dos.locales$t)
plot(dos.locales$t, localAtend, type = 'l', main='Ventas Locales
    A con tendencia', xlab='dias', ylab='ventas k$')
abline(tendencia, col='red')

localAdetend = localAtend - (tendencia$coefficients[1] +
    tendencia$coefficients[2]*dos.locales$t)

plot(dos.locales$t, dos.locales$localA, type =
    'l', main='Ventas Locales A', xlab='dias', ylab='ventas
    k$', ylim=c(min(localAdetend), max(dos.locales$localA)))
lines(dos.locales$t, localAdetend, col = 'red')
```

Detrending de ST con regresión



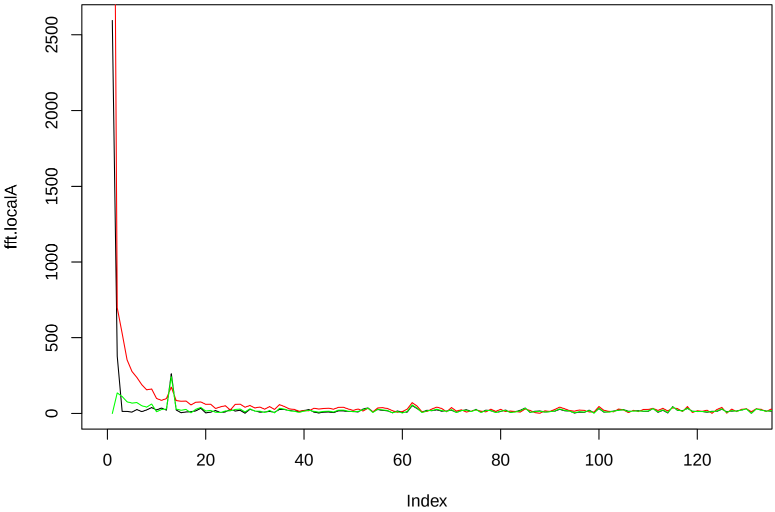
Detrending de ST con regresión



Detrending de ST con regresión

```
fft.localAdetend = Mod(fft(localAdetend))
plot(fft.localA,type='l',xlim=c(0,length(fft.localA)/2))
lines(fft.localAtend,col='red')
lines(fft.localAdetend,col='green')
```

Detrending de ST con regresión



Comparación de métodos de diagnóstico: método de Bland y Altman

- Antes del paper de J. Martin Bland y Douglas G. Altman, cuando se necesitaba comparar dos conjuntos de datos, apareados es decir de muestras repetidas, se utilizaba el análisis de correlación y la regresión, pero dichos métodos no son adecuados, se ha demostrado que no muestran las diferencias.
- Entonces en el paper **Martin Bland and Douglas Altman. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet , Volume 327 , Issue 8476 , 307 - 310**, los autores propusieron plotear la diferencia con respecto a la media entre cada una de las muestras apareadas.

paper de Bland y Altman

Measurement

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND   DOUGLAS G. ALTMAN

Department of Clinical Epidemiology and Social Medicine, St George's Hospital Medical School, London SW17; and Division of Medical Statistics, MRC Clinical Research Centre, Northwick Park Hospital, Harrow, Middlesex

Summary In clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they agree sufficiently for the new to replace the old. Such investigations are often analysed inappropriately, notably by using correlation coefficients. The use of correlation is misleading. An alternative approach, based on graphical techniques and simple calculations, is described, together with the relation between this analysis and the assessment of repeatability.

PEFR MEASURED WITH WRIGHT PEAK FLOW AND MINI WRIGHT PEAK FLOW METER

Subject	Wright peak flow meter		Mini Wright peak flow meter	
	First PEFR (l/min)	Second PEFR (l/min)	First PEFR (l/min)	Second PEFR (l/min)
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
5	476	470	500	500
6	557	611	600	625
7	413	415	364	460
8	442	431	380	390
9	650	638	658	642
10	433	429	445	432
11	417	420	432	420
12	656	633	626	605
13	267	275	260	227
14	478	492	477	467
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443

PLOTTING DATA

The first step is to plot the data and draw the line of equality on which all points would lie if the two meters gave exactly the same

Detrending de ST con regresión: aplicación del método de Bland y Altman

```
medialocalA = mean(dos.locales$localA)
print(medialocalA)
localA = dos.locales$localA - medialocalA

diferencia = localAdetend - localA
media = (localAdetend + localA)/2

plot(media, diferencia, ylim=c(-4,4))
abline(h=mean(diferencia))
abline(h=mean(diferencia)+2*sd(diferencia), lty=2)
abline(h=mean(diferencia)-2*sd(diferencia), lty=2)
abline(lm(diferencia ~ media), lty=3)
```

Detrending de ST con regresión: aplicación del método de Bland y Altman