Grados en Informática
Métodos Estadísticos para la Computación.
Control 7/8 Abril 2014 Tipo A

APELLIDOS, NOMBRE:
DNI: Titulación: Grupo:

1. (**MATLAB: Entregar en folio aparte.**) **Indicad solo las órdenes necesarias para resolver el problema.** Dada la tabla bidimensional de frecuencias absolutas:

$X\backslash Y$	-1	0	1	2	3
$(-\infty,-10]$	0	0	0	75	112
$(-10,0]$	0	0	73	505	5
$(0,5]$	143	121	63	0	0
$(5,15]$	230	71	8	0	0
$(15,\infty)$	50	6	0	0	0

Hallar:

(a) Las rectas de Y/X, X/Y y el coeficiente de correlación lineal.

(b) Ajustar una función de la forma: $\mathbf{y} = \frac{\mathbf{1}}{\mathbf{ax+bx^3}}$ y hallar la varianza residual correspondiente.

(c) Hallar la media y varianza de la variable $Y/(X \leq 0)$.

(1+0.75+0.75=2.5 Puntos)

2. Durante el estudio llevado a cabo durante varios años, de los factores que influyen en la contaminación del embalse "Las Moreras", se obtuvo la tabla de frecuencias bidimensional (X, Y):

$X\backslash Y$	$(-\infty,0]$	$(0,2]$	$(2,6]$	$(6,10]$
0	7	3	0	0
1	0	6	4	0
2	0	0	6	4

donde $X = 0$ corresponde a valores medidos el día 1 de mayo, $X = 1$ el 1 de junio y $X = 2$ el 1 de julio, mientras que, Y mide la materia seca (polución) en partes por millón.

(a) Usar el método de mínimos cuadrados para determinar el sistema de ecuaciones normales del modelo $\mathbf{Y = a + bX^2}$.

(b) Calcular la recta de regresión de $\mathbf{Y/X}$.

(c) Determinar el modelo que se ajusta mejor a los datos y úsalo para predecir el valor de la polución para el 1 de agosto ($X = 3$).

(d) Representar gráficamente y calcular media, moda y mediana de la variable Y condicionada a $X = 1$.

(1.25+1.25+1.25+1.25=5 Puntos)

3. El consumo en Tm. de aceite "El cencerro" se ha medido durante varios años resultando los siguientes valores.

	2010			2011			2012		
	I	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*
Consumo	2.2	2.3	3.1	3.4	3.3	3.8	4.4	4.6	5.2

donde I, II y III indica el cuatrimestre correspondiente.

Depurar la tabla eliminando la influencia de la tendencia y del cuatrimestre mediante el uso de medias móviles de orden 3.

(1+1+0.5=2.5 Puntos)

Grados en Informática
Métodos Estadísticos para la Computación.
Control 7/8 Abril 2014   Tipo B

APELLIDOS, NOMBRE:
DNI: Titulación: Grupo:

---

1. (**MATLAB: Entregar en folio aparte.**) **Indicad solo las órdenes necesarias para resolver el problema.** Dada la tabla bidimensional de frecuencias absolutas:

$X\backslash Y$	0	1	2	3	4
$[0, 10]$	0	0	0	77	222
$(10, 20]$	0	0	73	555	51
$(20, 25]$	163	125	65	0	0
$(25, 30]$	130	50	0	0	0
$(30, \infty)$	56	6	0	0	0

Hallar:

(a) Las rectas de Y/X, X/Y y el coeficiente de correlación lineal.

(b) Ajustar una función de la forma: $\mathbf{y} = \frac{\mathbf{1}}{\mathbf{a+bLn(x)}}$ y hallar la varianza residual correspondiente.

(c) Hallar media, varianza y moda de la marginal de la variable Y.

(1+0.75+0.75=2.5 Puntos)

2. Durante el estudio llevado a cabo durante varios años, de los factores que influyen en el crecimiento del girasol, se obtuvo la tabla de frecuencias bidimensional (X, Y):

$X\backslash Y$	$[0, 3]$	$(3, 5]$	$(5, \infty]$
0	5	3	0
1	4	6	2
2	0	5	5

donde $X = 0$ indica que no se ha abonado , $X = 1$ que se ha abonado con 100Kg. por Ha., $X = 2$ que lo ha sido a 200Kg. por Ha., Y mide el rendimiento en Tm/Ha.

(a) Usar el método de mínimos cuadrados para ajustar una función del tipo $\mathbf{Y} = \mathbf{a} + \frac{\mathbf{b}}{\mathbf{1+X^2}}$.

(b) Calcular la recta de regresión de $\mathbf{Y/X}$.

(c) Determinar el modelo que se ajusta mejor a los datos y úsalo para predecir el rendimiento cuando X=1.5 (150 Kg. por Ha).

(d) Representar gráficamente y calcular media, mediana y desviación media de la variable Y condicionada a $X = 1$.

(1.25+1.25+1.25+1.25=5 Puntos)

3. La Finca "La Flauta" produce patatas (X) y tomates (Y), resultando los siguientes valores para la producción en Tm y precio de venta cada 10 Kg.

	2010		2011		2012		2013	
	Patatas	*Tomates*	*Patatas*	*Tomates*	*Patatas*	*Tomates*	*Patatas*	*Tomates*
Producción	23	11	25	13	27	14	29	15
Precio venta	2	5	2	6	3	7	3	8

Calcular los índices de precios de Laspeyres, Paasche y Fisher de los años 2012 y 2013 con base el año 2010.

(1+1+0.5=2.5 Puntos)

Grados en Informática

Métodos Estadísticos para la Computación.

Control 7/8 Abril 2014 Tipo C

APELLIDOS, NOMBRE:
DNI: Titulación: Grupo:

---

1. (**MATLAB: Entregar en folio aparte.**) **Indicad solo las órdenes necesarias para resolver el problema.** Dada la tabla bidimensional de frecuencias absolutas:

$X\backslash Y$	$(-\infty, 0]$	$(0, 10]$	$(10, 20]$	$(20, 25]$	$(25, \infty)$
$(-\infty, -10]$	0	0	0	115	212
$(-10, 0]$	0	0	78	515	51
$(0, 5]$	140	140	66	0	0
$(5, 15]$	230	75	3	0	0
$(15, \infty)$	55	10	0	0	0

Hallar:

(a) Las rectas de Y/X, X/Y y el coeficiente de correlación lineal.

(b) Ajustar una función de la forma: $\mathbf{y = ax + bsen(x)}$ y hallar la varianza residual correspondiente.

(c) Hallar la media y varianza de la variable $Y/(X > 0)$.

(1+0.75+0.75=2.5 Puntos)

2. El tiempo de respuesta de un terminal en un sistema (T) es función del número de usuarios conectados (X). Para estimar dicha función se miden los tiempos de respuesta en determinadas circunstancias, obteniéndose la tabla:

X	0	1	2	5	10
T	1.5	2	2.7	4	7

(a) Usar el método de mínimos cuadrados para ajustar una recta de T/X.

(b) Ajustar una función del tipo $\mathbf{T = 1 + e^{a+bX}}$.

(c) Determinar el modelo que se ajusta mejor a los datos.

(d) Estimar, mediante el mejor modelo, el tiempo que tardará el terminal cuando tenga conectados 20 usuarios.

(1.25+1.25+1.25=3.75 Puntos)

3. Se ha medido el consumo de agua (en miles de m^3) y el costo del m^3 (en céntimos de euro) del regadío de la Finca "La Flauta", resultando los siguientes valores:

	2009		2010		2011		2012	
	I	*II*	*I*	*II*	*I*	*II*	*I*	*II*
Consumo	23	19	25	17	27	20	27	20
Costo	2	4	2	4.3	2.1	4.9	2.2	5.0

donde I corresponde al semestre desde {Octubre(año anterior) a Marzo} y II al de {Abril a Septiembre}.

Calcular:

(a) Los índices de variación estacional para el consumo semestral y **desestacionalizar** la serie original. Interpreta los resultados e indica el año y semestre con consumo anormalmente grande si lo hubiese.

(b) Agrupa los valores del costo y consumo por años y calcula (base 2009) el índice simple y de cadena del año 2012 para el consumo y para el costo. NOTA: El consumo será el consumo total por año y el costo medio del m^3, la media de los costos semestrales ponderados con el consumo.

((1.25+0.5)+(0.75+0.75)=3.25 Puntos)

Soluciones: Examen A

Problema 1:

```
clc,clear all,format compact
disp('Problema 1')
disp('1-a')
x=[-15 -15 -5 -5 -5 2.5 2.5 2.5 10 10 10 20 20]
y=[2 3 1 2 3 -1 0 1 -1 0 1 -1 0]
n=[75 112 73 505 5 143 121 63 230 71 8 50 6]
N=sum(n)
disp('Recta Y/X')
A1=[N sum(n.*x);sum(n.*x) sum(n.*x.^2)]
b1=[sum(n.*y); sum(n.*y.*x)]
sol=A1\b1
a=sol(1), b=sol(2)
disp('La recta es y=a+bx')

disp('Recta X/Y')
A2=[N sum(n.*y);sum(n.*y) sum(n.*y.^2)]
b2=[sum(n.*x); sum(n.*y.*x)]
sol2=A2\b2
c=sol2(1), d=sol2(2)
disp('La recta es x=c+dy')
r=sign(b)*sqrt(b*d)

disp(1-b:')

disp('Otro ajuste')
% Hacemos el cambio Y=1/y=ax+bx^3 quedando:
% <Y,x>=a<x,x>+b<x,x^3>
% <Y,x^3>=a<x,x^3>+b<x^3,x^3>
Y=1./y
A3=[sum(n.*x.^2) sum(n.*x.^4);sum(n.*x.^4) sum(n.*x.^6)]
b3=[sum(n.*y.*x);sum(n.*y.*x.^3)]
sol3=A3\b3
p=sol3(1),q=sol3(2)
disp('La solución es y=1/(px+qx^3)')
yest=1./(p*x+q*x.^3)
e=y-yest
sne=sum(n.*e)
sne2=sum(n.*e.^2)
Vr=sne2/N-(sne/N)^2

disp(1-c:')
yc=[1 2 3] % Condicionada
nc=[73 75+505 112+5]
mediac=sum(nc.*yc)/sum(nc)
m2c=sum(nc.*yc.^2)/sum(nc)
Varc=m2c-mediac^2
```

Problema 2:
a:

Y	x_i	y_i	n_i	n_ix_i	n_iy_i	$n_ix_i^2$	$n_iy_i^2$	$n_ix_iy_i$	$n_iy_ix_i^2$	$n_ix_i^4$	y_i^{est}	$e_i\!=\!y_i\!-\!y_i^{est}$	n_ie_i	$n_ie_i^2$
$(-2,0]$	0	-1	7	0	-7	0	7	0	0	0	0.1077	-1.1077	-7.7538	8.5889
$(0,2]$	0	1	3	0	3	0	3	0	0	0	0.1077	0.8923	2.6769	2.3886
$(0,2]$	1	1	6	6	6	6	6	6	6	6	1.5231	-0.5231	-3.1385	1.6417
$(2,6]$	1	4	4	4	16	4	64	16	16	4	1.5231	2.4769	9.9077	24.5406
$(2,6]$	2	4	6	12	24	24	96	48	96	96	5.7692	-1.7692	-10.6154	18.7811
$(6,10]$	2	8	4	8	32	16	256	64	128	64	5.7692	2.2308	8.9231	19.9053
			30	30	74	50	432	134	246	170			0	75.8462

Hacemos el cambio $X = x^2$ quedando: $y = a + bX$, por lo que las ecuaciones normales son:

$$\left.\begin{array}{ll} \sum_i y_i &= Na \quad +b\sum_i X_i \\ \sum_i y_iX_i &= a\sum_i X_i \quad +b\sum_i X_i^2 \end{array}\right\} \Rightarrow \left.\begin{array}{ll} \sum_i y_i &= Na \quad +b\sum_i x_i^2 \\ \sum_i y_ix_i^2 &= a\sum_i x_i^2 \quad +b\sum_i x_i^4 \end{array}\right\} \Rightarrow \left.\begin{array}{lll} 74 &= 30a & +50b \\ 246 &= 50a & +170b \end{array}\right\}$$

$\Rightarrow a = 0.1077,\ \ b = 1.4154 \Rightarrow \mathbf{y = 0.1077 + 1.4154x^2}$

b:

$$\left.\begin{array}{ll} \sum_i y_i &= Na \quad +b\sum_i x_i \\ \sum_i y_ix_i &= a\sum_i x_i \quad +b\sum_i x_i^2 \end{array}\right\} \Rightarrow \left.\begin{array}{lll} 74 &= 30a & +30b \\ 134 &= 30a & +50b \end{array}\right\}$$

$\Rightarrow a = -0.5333,\ \ b = 3.0000 \Rightarrow \mathbf{y = -0.5333 + 3x}$

c:
Para el ajuste no lineal:
Tras calcular las columnas de la tabla $y_i^{est} = 0.1077 + 1.4154x_i^2$, $e_i = y_i - y_i^{est}$, n_ie_i y $n_ie_i^2$ calculamos la varianza residual: $\mathbf{V_r^a} = \frac{\sum_i n_ie_i^2}{N} - \left(\frac{\sum_i n_ix_i}{N}\right)^2 = \frac{75.8462}{30} - 0^2 = \mathbf{2.5282} \Rightarrow$

$\mathbf{R^2} = 1 - \frac{V_r^a}{V_y} = 1 - \frac{2.5282}{8.3156} = \mathbf{0.6960}$

Para la recta:

$V_x = \frac{\sum_i n_ix_i^2}{N} - \left(\frac{\sum_i n_ix_i}{N}\right)^2 = \frac{50}{30} - \left(\frac{30}{30}\right)^2 = 0.6667 \Rightarrow \sigma_x = 0.8165$

$cov = \frac{\sum_i n_ix_iy_i}{N} - \frac{\sum_i n_ix_i}{N}\frac{\sum_i n_iy_i}{N} = \frac{134}{30} - 1\frac{74}{30} = 2$

$\mathbf{r} = \frac{\mathbf{cov}}{\sigma_\mathbf{x}\sigma_\mathbf{y}} = \frac{\mathbf{2}}{\mathbf{(0.8165)(2.8837)}} = \mathbf{0.8494}, \qquad \mathbf{r^2 = 0.7215}$

$\mathbf{V_r^b = V_y(1 - r^2) = 2.3156}$

De lo que se deduce que **es mejor el ajuste lineal** pues la varianza residual es menor, o bien, $r^2 > R^2$.
Para $x = 3$, $\mathbf{y(3) = -0.5333 + 3(3) = 8.4667}$

d:
Al no existir individuos con $x = 1$ en las clases $(\infty, 0]$ y $(6, 10]$, la variable $Y/(X = 1)$ queda solo con 2 modalidades:

$Y/(X=1)$	n_i	a_i	h_i	y_i^c	$n_iy_i^c$	N_i
$[0,2]$	6	2	3	1	6	6
$(2,6]$	4	4	1	4	16	10
	10				22	

La representación gráfica apropiada para un carácter continuo es el histograma.

Media: $\mathbf{\bar{Y/(X = 1)} = \frac{22}{10} = 2.2}$
Moda: El intervalo modal es el $[0, 2]$ pues su $h_i = 3$ y es el mayor: $\Delta_1 = 3 - 0 = 3$, $\Delta_2 = 3 - 1 = 2$ y $\mathbf{Mo = 0 + \frac{3}{3+2} * 2 = 1.2}$
Mediana: $N/2 = 5$ y el intervalo mediano es el $[0, 2]$, siendo la mediana: $\mathbf{Me = 0 + \frac{5-0}{6} * 2 = 1.6667}$

Problema 3:
Disponemos los datos X en forma de tabla, y la tendencia (medias móviles orden 3) en la tabla $\hat{T}_3$:

X	2010	2011	2012
I	2.2	3.4	4.4
II	2.3	3.3	4.6
III	3.1	3.8	5.2

,

$\hat{T}_3$	2010	2011	2012
I	$-$	3.2667	4.2667
II	2.5333	3.5000	4.7333
III	2.9333	3.8333	$-$

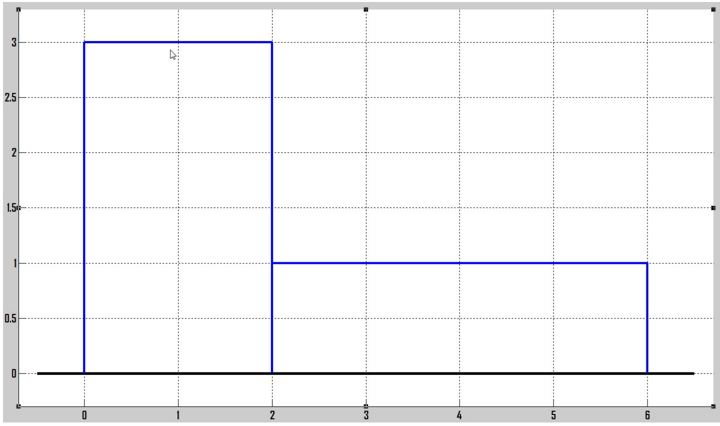

Figure 1: Histograma de frecuencias.

Eliminamos el factor tendencia (tabla X/T) y calculamos los índices de estacionalidad sin corregir:

X/T	2010	2011	2012
I	$-$	1.0408	1.0313
II	0.9079	0.9429	0.9718
III	1.0568	0.9913	$-$

$$\Rightarrow \begin{array}{l} I_I = \frac{1.0408+1.0313}{2} = 1.0360 \\ I_{II} = \frac{0.9079+0.942+0.9718}{3} = 0.9409 \\ I_{III} = \frac{1.0568+0.9913}{2} = 1.0241 \end{array}$$

Calculamos su suma $S = I_I + I_{II} + I_{III} = 3.0010$. Y, aunque el valor obtenido es próximo a 3 y en este caso puede evitarse este paso, corregimos:

$\mathbf{I^c_I = I_I * \frac{3}{3.001} = 1.0357}$, $\mathbf{I^c_{II} = I_{II} * \frac{3}{3.001} = 0.9406}$, $\mathbf{I^c_{III} = I_{III} * \frac{3}{3.001} = 1.0237}$

Por último y contestando a lo pedido en el problema, obtenemos la tabla en donde se han eliminado, tanto la tendencia, como la estacionalidad, mediante la división de la X/T por el índice corregido correspondiente, quedando solo los factores accidentales:

A	**2010**	**2011**	**2012**
I	$-$	**1.0049**	**0.9957**
II	**0.9653**	**1.0024**	**1.0332**
III	**1.0323**	**0.9683**	$-$

Examen B

Problema 1:

```
clc,clear all,format compact
disp('1-a')
x=[5 5 15 15 15 22.5 22.5 22.5 27.5 27.5 32.5 32.5]
y=[3 4 2 3 4 0 1 2 0 1 0 1]
n=[77 222 73 555 51 163 125 65 130 50 56 6]
N=sum(n)
A1=[N sum(n.*x)
   sum(n.*x) sum(n.*x.^2)]
b1=[sum(n.*y); sum(n.*y.*x)]
sol1=A1\b1,a=sol1(1), b=sol1(2)
% La recta Y/X es y=a+bx
A2=[N sum(n.*y)
   sum(n.*y) sum(n.*y.^2)]
b2=[sum(n.*x); sum(n.*y.*x)]
sol2=A2\b2,ap=sol2(1), bp=sol2(2)
% La recta X/Y es x = ap + bp y
r=sqrt(b*bp); if b<0,r=-r;end,r
disp('1-b')
%Hago el cambio Y=1/y, X=Ln(x)
Y=1./y
X=log(x)
A3=[N sum(n.*X)
   sum(n.*X) sum(n.*X.^2)]
b3=[sum(n.*Y);sum(n.*Y.*X)]
sol3=A3\b3
a=sol3(1),b=sol3(2)
% La funcion ajustada es y=1/(a+bLn(x))
yest=1./(a+b*log(x))
e=y-yest
Vr=sum(n.*e.^2)/N-(sum(e)/N)^2
disp('1-c')
ym=0:4
nm=[163+130+56 125+50+6 73+65 77+555 222+51]
medY=sum(ym.*nm)/N
varY=sum(ym.^2.*nm)/N-medY^2
```

NOTA: Al ejecutar, para estos datos concretos, daría problemas el apartado 1-b.

Problema 2:

Los cálculos de todo el problema los hacemos con la tabla:

X_i	Y_i	n_i	$x_i = \frac{1}{1+X_i^2}$	$\frac{n_i}{1+X_i^2}$	$\frac{n_i}{(1+X_i^2)^2}$	n_iX_i	$n_iX_i^2$	$n_iX_iY_i$	n_iY_i	$n_iY_i^2$	$\frac{n_iY_i}{1+X_i^2}$
0	1.5	5	1	5	5	0	0	0	7.5	11.25	7.5
0	4	3	1	3	3	0	0	0	12	48	12
1	1.5	4	0.5	2	1	4	4	6	6	9	3
1	4	6	0.5	3	1.5	6	6	24	24	96	12
1	6	2	0.5	1	0.5	2	2	12	12	72	6
2	4	5	0.2	1	0.2	10	20	40	20	80	4
2	6	5	0.2	1	0.2	10	20	60	30	180	6
		30		16	11.4	32	52	142	111.5	496.25	50.5

a: Hacemos el cambio de variable $x = \frac{1}{1+X^2}$ quedando $Y = a + bx$

Las ecuaciones normales son:

$$\left.\begin{array}{lll}\sum Y_i & = Na & +b\sum x_i \\ \sum Y_i x_i & = a\sum x_i & +b\sum x_i^2\end{array}\right\} \Rightarrow \left.\begin{array}{lll}111.5 & = 30a & +16b \\ 50.5 & = 16a & +11.4b\end{array}\right\} \Rightarrow \begin{array}{c} a = 5.3849 \\ b = -3.1279\end{array}$$

Por tanto el ajuste pedido es:

$$\boldsymbol{Y = 5.3849 + \frac{-3.1279}{1+X^2}}$$

b: Ahora las ecuaciones normales son:

$$\left.\begin{array}{lll}\sum Y_i & = Na & +b\sum X_i \\ \sum Y_i X_i & = a\sum X_i & +b\sum X_i^2\end{array}\right\} \Rightarrow \left.\begin{array}{lll}111.5 & = 30a & +32b \\ 142 & = 32a & +52b\end{array}\right\} \Rightarrow \begin{array}{c} a = 5.3849 \\ b = -3.1279\end{array}$$

Por tanto el ajuste pedido es:

$$\boldsymbol{Y = 2.3396 + 1.2910X}$$

c: El mejor ajuste puede determinarse comparando las varianzas residuales, o bien, los coeficientes de determinación R^2.
Caso lineal: Calculamos el coeficiente de correlación lineal: $r = \frac{cov}{\sigma_X \sigma_Y}$
$V_X = \frac{52}{30} - \left(\frac{32}{30}\right)^2 = 0.5956 \Rightarrow \sigma_X = \sqrt{0.5956} \approx 0.7717$
$V_Y = \frac{496.25}{30} - \left(\frac{111.5}{30}\right)^2 = 2.7281 \Rightarrow \sigma_Y = \sqrt{2.7281} \approx 1.6517$
$Cov(X,Y) = \frac{142}{30} - (\frac{32}{30}\frac{111.5}{30}) \approx 0.7689 \Rightarrow \mathbf{r} = \frac{\mathbf{0.7689}}{\mathbf{0.7717*1.6517}} = \mathbf{0.6032}$

$$\boldsymbol{r^2 \approx 0.3639, \qquad V_r^b = V_Y(1-r^2) \approx 1.7354}$$

Caso no lineal: Calculamos los valores estimados por el ajuste, errores y varianza de ellos. De ahí obtenemos el coeficiente de determinación mediante $R^2 = 1 - \frac{V_r}{V_Y}$.

X_i	Y_i	n_i	$Y_i^{est} = 5.3849 + \frac{-3.1279}{1+X_i^2}$	$e_i = Y_i - Y_i^{est}$	$n_i e_i$	$n_i e_i^2$
0	1.5	5	2.2570	−0.7570	3.7849	2.8651
0	4.0	3	2.2570	1.7430	5.2291	9.1144
1	1.5	4	3.8209	−2.3209	−9.2837	21.5469
1	4.0	6	3.8209	0.1791	1.0744	0.1924
1	6.0	2	3.8209	2.1791	4.3581	9.4967
2	4.0	5	4.7593	−0.7593	−3.7965	2.8827
2	6.0	5	4.7593	1.2407	6.2035	7.6967
					0	53.7948

$\mathbf{V_r^a} = \frac{53.7948}{30} - (\frac{0}{30})^2 = \mathbf{1.7932 \Rightarrow R^2 = 1 - \frac{1.7932}{2.7281} \approx 0.3427}$, luego **es ligeramente mejor el ajuste lineal**.
Para X=1.5, sustituimos en el ajuste lineal: $Y(1.5) = Y = 2.3396 + 1.2910 * 1.5 = \mathbf{4.2761}$

d: La condicionada de Y a X=1 es:

$Y/_{X=1}$	n_i	y_i	$n_i y_i$	$n_i y_i^2$	a_i	h_i	N_i	d_i	$n_i d_i$
$[0,3]$	4	1.5	6	9	3	1.333	4	2	8
$(3,5]$	6	4	24	96	2	3.000	10	0.5	3
$(5,\infty)$	2	6	12	72	2	1.000	12	2.5	5
	12		42	177					16

Media: $\bar{\mathbf{x}} = \frac{\mathbf{42}}{\mathbf{12}} = \mathbf{3.5}$,

Desviación media: Calculo las desviaciones (respecto a la media) $d_i = |y_i - 3.5|$ y su media

$$\mathbf{DM} = \frac{\sum_i \mathbf{d_i}}{\mathbf{12}} = \frac{\mathbf{16}}{\mathbf{12}} \approx \mathbf{1.3333}$$

Mediana: $N/2 = 6$, en N_i la primera que lo rebasa es la clase $(3, 5]$, así:

$$\mathbf{Me} = \mathbf{3} + \frac{\mathbf{6} - \mathbf{4}}{\mathbf{6}}\mathbf{2} \approx \mathbf{3.6667}$$

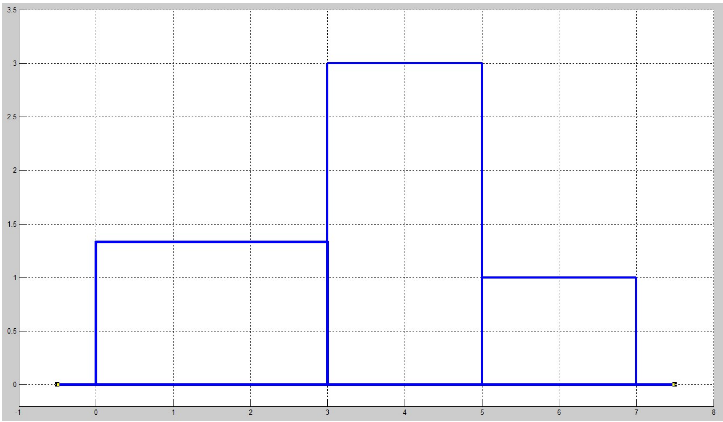

Figure 2: Histograma de frecuencias.

Problema 3:
Laspeyres:

$$L_{2012/2010} = \frac{3 * 23 + 7 * 11}{2 * 23 + 5 * 11} = 1.4455, \qquad L_{2013/2010} = \frac{3 * 23 + 8 * 11}{2 * 23 + 5 * 11} = 1.5545$$

Paashe:

$$P_{2012/2010} = \frac{3 * 27 + 7 * 14}{2 * 27 + 5 * 14} = 1.4435, \qquad P_{2013/2010} = \frac{3 * 29 + 8 * 15}{2 * 29 + 5 * 15} = 1.5564$$

Fisher:

$$F_{2012/2010} = \sqrt{1.4455 * 1.4435} = 1.4445, \qquad F_{2013/2010} = \sqrt{1.5545 * 1.5564} = 1.5554$$

Examen C
Problema 1:

```
clc,clear all,format compact
disp('1-a')
x=[-15 -15 -5 -5 -5 2.5 2.5  2.5 10 10 10 20 20]
y=[22.5 27.5 15 22.5 27.5 -5 5 15 -5 5 15 -5 15]
n=[115 212 78 515 51 140 140 66 230 75 3 55 10]
N=sum(n)
A1=[N sum(n.*x)
   sum(n.*x) sum(n.*x.^2)]
b1=[sum(n.*y); sum(n.*y.*x)]
sol1=A1\b1,a=sol1(1), b=sol1(2)
% La recta Y/X es y=a+bx
A2=[N sum(n.*y)
   sum(n.*y) sum(n.*y.^2)]
b2=[sum(n.*x); sum(n.*y.*x)]
sol2=A2\b2,ap=sol2(1), bp=sol2(2)
% La recta X/Y es x = ap + bp y
r=sqrt(b*bp); if b<0,r=-r;end,r
disp('1-b')
A3=[sum(n.*x.^2) sum(n.*x.*sin(x))
   sum(n.*x.*sin(x)) sum(n.*sin(x).^2)]
b3=[sum(n.*y.*x);sum(n.*y.*sin(x))]
sol3=A3\b3
a=sol3(1),b=sol3(2)
% La funcion ajustada es y=ax+bsen(x)
yest=a*x+b*sin(x)
e=y-yest
Vr=sum(n.*e.^2)/N-(sum(e)/N)^2
disp('1-c')
Yc=[-5 5 15]
nc=[140+230+55 140+75+10 66+3]
Nc=sum(nc)
medYc=sum(Yc.*nc)/Nc
varYc=sum(Yc.^2.*nc)/Nc-medYc^2
```

Problema 2
a: Los cálculos, de todo el problema, vamos a ponerlos en la tabla siguiente:

X_i	T_i	X_i^2	X_iT_i	$Y_i = Ln(T_i - 1)$	Y_iX_i	T_i^*	$e_i = T_i - T_i^*$	e_i^2	T_i^2
0	1.5	0	0	-0.6931	0	1.7684	-0.2684	0.0721	2.25
1	2	1	2	0	0	1.9620	0.0380	0.0014	4
2	2.7	4	5.4	0.5306	1.0613	2.2044	0.4956	0.2456	7.29
5	4	25	20	1.0986	5.4931	3.3636	0.6364	0.4050	16
10	7	100	70	1.7818	17.9176	8.2701	-1.2701	1.6131	49
18	17.2	130	97.4	2.7279	24.4719	17.5686	-0.3686	2.3372	78.54

Así, $\bar{X} = \frac{18}{5} = 3.6$, $\bar{T} = \frac{12.2}{5} = 3.44$, $V_X = \frac{130}{5} - 3.6^2 = 13.04$,
$Cov(X,T) = \frac{97.4}{5} - 3.6(3.44) = 7.096$ y la recta de T/X queda: $T - 3.44 = \frac{7.096}{13.04}(X - 3.6) \Rightarrow$

$$\mathbf{T = 1.4810 + 0.5442X}$$

b: Buscamos el cambio a realizar: $T-1=e^{a+bx}$, $Ln(T-1)=a+bX$. Luego el cambio es Y=Ln(T-1), quedando Y=a+bX.

Las ecuaciones normales son:

$$\left.\begin{array}{lll}\sum Y_i & =Na & +b\sum X_i \\ \sum X_iY_i & =a\sum X_i & +b\sum X_i^2\end{array}\right\} \Rightarrow \left.\begin{array}{lll}2.7279 & =5a & +18b \\ 24.4719 & =18a & +130b\end{array}\right\} \Rightarrow a=-0.2634, \quad b=0.2247$$

Por tanto:

$$\mathbf{T=1+e^{-0.2634+0.2247X}}$$

c: Podemos basarnos en r^2 (o para el segundo ajuste en R^2), o bien, en las varianzas residuales.

Calculamos la varianza de T: $V_T=\frac{78.54}{5}-3.44^2=3.8744$

Para la recta, calculamos $r=\frac{cov}{\sigma_X\sigma_T}=\frac{7.096}{\sqrt{13.04}\sqrt{3.8744}}=0.9983 \Rightarrow r^2=0.9967 \Rightarrow$

$$\mathbf{V_r^{recta}=V_T(1-r^2)=3.8744(1-0.9967)\approx 0.013}$$

Para el segundo ajuste, debemos calcular los valores estimados: $T_i^*=1+e^{-0.2634+0.2247X_i}$, los errores $e_i=T_i-T_i^*$ y la varianza de estos:

$$V_r^{II}=\frac{\sum_i e_i^2}{N}-\left(\frac{\sum_i e_i}{N}\right)^2=\frac{2.3372}{5}-\left(\frac{-0.3686}{5}\right)^2=0.4620$$

Luego **es mejor ajuste el lineal,** ya que la varianza residual es menor.

También podríamos llegar a esa conclusión comparando R^2 y r^2, el que sea mayor es mejor, y aquí se da: $r^2>R^2$, pues $R^2=1-0.462/3.8744=0.8808$ y, ya se obtuvo $r^2=0.9967$

d: Sustituimos X=20 en la recta: $T=1.4810+0.5442(20)=\mathbf{12.3644}$

Problema 3:

a: Ponemos las serie temporal en forma de tabla:

X	2009	2010	2011	2012
I	23	25	27	27
II	19	17	20	20

Al existir 2 periodos (semestres) al año, usamos medias móviles de orden 2.

T'_2	2009	2010	2011	2012
I				
	21	21	23.5	23.5
II				
	22	22	23.5	

Ahora centramos los valores de la tendencia en el tiempo:

X	2009	2010	2011	2012
I	$-$	21.5	22.75	23.5
II	21.5	21.5	23.5	$-$

Eliminamos la tendencia:

X/T	2009	2010	2011	2012
I	$-$	1.1628	1.1868	1.1489
II	0.8837	0.7907	0.8511	$-$

y calculamos los índices de variación estacional (sin corregir): $IVE(I)=\frac{1.1628+1.1868+1.1489}{3}=1.1662$, $IVE(II)=\frac{0.8837+0.7907+0.8511}{3}=0.8418$

Calculamos su suma: $S = IVE(I) + IVE(II) = 1.1662 + 0.8418 = 2.008$ y los corregimos: $IVE_c(I) = IVE(I) * 2/S = 1.1615$, $IVE_c(II) = IVE(II) * 2/S = 0.8385$

Para desestacionalizar la serie temporal dividimos la serie original entre los índices corregidos:

$\boldsymbol{X/E}$	**2009**	**2010**	**2011**	**2012**
$\boldsymbol{I}$	**19.8015**	**21.5233**	**23.2452**	**23.2452**
$\boldsymbol{II}$	**22.6595**	**20.2750**	**23.8530**	**23.8530**

b: Agrupamos los valores:

Año	2009	2010	2011	2012
Consumo	42	42	47	47
Costo	2.9048	2.9310	3.2915	3.3915

donde Costo(2009) $= \frac{2*23+4*19}{23+19} = 2.9048$, Costo(2010) $= \frac{2*25+4.3*17}{25+17} = 2.9310$, Costo(2011) $= \frac{2.1*27+4.9*20}{27+20} = 3.2915$ y Costo(2012) $= \frac{2.2*27+5*20}{27+20} = 3.3915$.

Para el Consumo:

Índice simple: $I_{12/09} = 47/42 = 1.119048 \Rightarrow 111.9048\%$

Índice cadena: $I'_{12/11} = 47/47 = 1 \Rightarrow 100\%$

Para el Costo:

Índice simple: $I_{12/09} = 3.3915/2.9048 = 1.167550 \Rightarrow 116.7550\%$

Índice cadena: $I'_{12/11} = 3.3915/2.2915 = 1.030381 \Rightarrow 103.0381\%$