PROBLEMAS DE ESTADÍSTICA INDUSTRIAL
(Ejercicios de estadística para Ingenieros)

Rosa Rodríguez Huertas, Antonio Gámez Mellado,
Luis Marín Trechera y Santiago Fandiño Patiño

Escuela Superior de Ingeniería.
Universidad de Cádiz

Diciembre de 2005

Índice General

Unidad Temática I

PROBLEMAS DE ESTADISTICA BÁSICA

T. 1

Estadística descriptiva

Ejercicio 1 *Una empresa de alimentación se dedica a enviar pizzas a domicilio. El número de pizzas enviadas en cada uno de los 30 días del mes de Abril son:*

47 63 66 58 32 61 57 44 44 56
38 35 76 58 48 59 67 33 69 53
51 28 25 36 49 78 48 42 72 52

1. *Construir una tabla de frecuencia relativa y de frecuencia acumulada usando una tabla tipo III e intervalos de clase con una amplitud de 5 pizzas.*

2. *Calcular la media en los siguientes casos: a)Usando todos los datos, b)Usando solamente los valores de la tabla de frecuencias construida*

3. *Calcular la, mediana, moda y la desviación típica usando los valores de la tabla*

4. *Determinar los tres cuartiles a partir de la tabla de frecuencias*

5. *Construir el histogramas de frecuencias y el polígono de frecuencias correspondiente.*

1. La tabla incluye, además de las frecuencias absolutas y relativas tanto simples y acumuladas, algunos cálculos para facilitar el cálculo de los

parámetros estadísticos que se requieren.

Intervalos de clase	x_i	$\mathbf{n}_i$	$\mathbf{N}_i$	$\mathbf{f}_i$	%	$\mathbf{F}_i$	$\mathbf{n}_i x_i$	$\mathbf{n}_i {x_i}^2$
[25,30)	27.5	2	2	$\frac{2}{30}$	6.67%	$\frac{2}{30}$	55.0	1512.5
[30,35)	32.5	2	4	$\frac{2}{30}$	6.67%	$\frac{4}{30}$	65.0	2112.5
[35,40)	37.5	3	7	$\frac{3}{30}$	10%	$\frac{7}{30}$	112.5	4218.75
[40,45)	42.5	3	10	$\frac{3}{30}$	10%	$\frac{10}{30}$	127.5	5418.75
[45,50)	47.5	4	14	$\frac{4}{30}$	13.3%	$\frac{14}{30}$	190.0	9025
[50,55)	52.5	3	17	$\frac{3}{30}$	10%	$\frac{17}{30}$	157.5	8268.75
[55,60)	57.5	5	22	$\frac{5}{30}$	16.67%	$\frac{22}{30}$	287.5	16531.25
[60,65)	62.5	2	24	$\frac{2}{30}$	6.67%	$\frac{24}{30}$	125.0	7812.5
65,70)	67.5	3	27	$\frac{3}{30}$	19%	$\frac{27}{30}$	202.5	13668.75
[70,75)	72.5	1	28	$\frac{1}{30}$	3.3%	$\frac{28}{30}$	72.5	5256.25
[75,80)	77.5	2	30	$\frac{2}{30}$	6.67%	1	155.0	12012.5
Totales		30		1	100%		1550	85838

2. Si tenemos en cuenta los datos primitivos hay que sumar todos ellos y dividir por el número de elementos de la muestra (30):

$$\frac{47+63+66+\cdots+42+72+52}{30}=\frac{1545}{30}=51.5$$

Si no disponemos los datos primitivos, pero si de la tabla de frecuencia la media se haría usando como datos las marcas de clase. En este caso la media tomaría el valor:

$$\frac{27.5\times 2+32.5\times 2+\cdots+72.5\times 1+77.5\times 2}{30}=\frac{1550}{30}=51.667$$

3. Para calcular la mediana calculamos previamente el intervalo mediano, que resulta ser [50,55). Suponiendo que los valores de la vatiable se distribuyen uniformemente en el intervalo mediano, la mediana tomaría el valor:

$$M_e=50+\frac{15-14}{3}\times 5=51.667$$

Para hallar la moda calculamos el intervalo modal. Como los intervalos son de igual amplitud este intervalo es el de mayor frecuencia [55,60). La moda se calculará con la expresión:

$$\begin{aligned} Mo &= L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i \\ \Delta_1 &= h_i - h_{i-1}, \quad \Delta_2 = h_i - h_{i+1}, \end{aligned}$$

$$Mo = 55 + \frac{2}{3+2} \times 5 = 57.0$$

La desviación típica, usando los valores agrupados y la expresión de la varianza que consiste en hallar la media de los cuadrados menos el cuadrado de la media sería:

$s = \sqrt{\frac{27.5^2 \times 2 + 32.5^2 \times 2 + \cdots + 72.5^2 \times 1 + 77.5^2 \times 2}{30} - 51.667^2} =$

$= \sqrt{\frac{1512.5 + 2112.5 + \cdots + 5256.3 + 12013.}{30} - 51.667^2} =$

$= \sqrt{\frac{85838.}{30} - 2669.5} = 13.848$

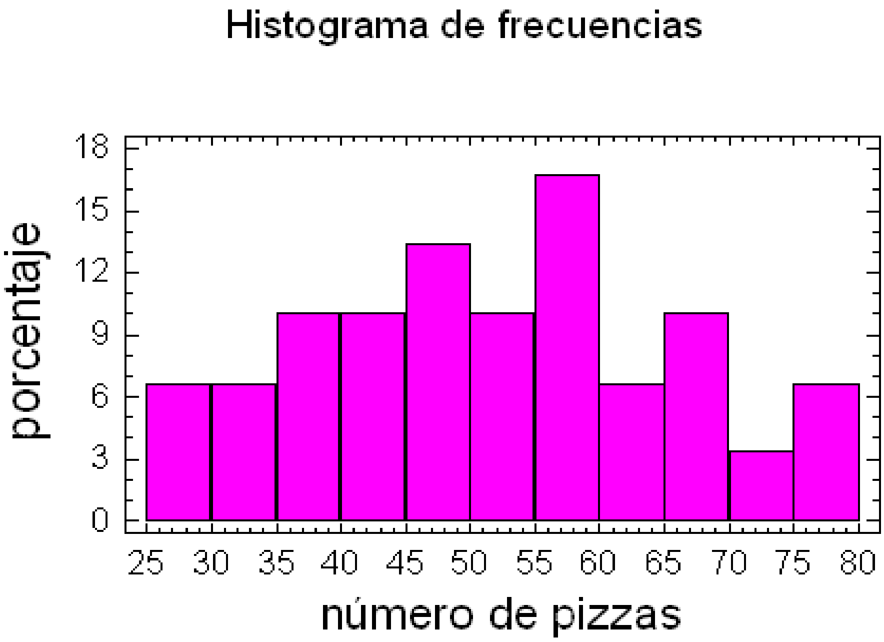
4. Para calcular los cuartiles, utilizamos las fórmulas.

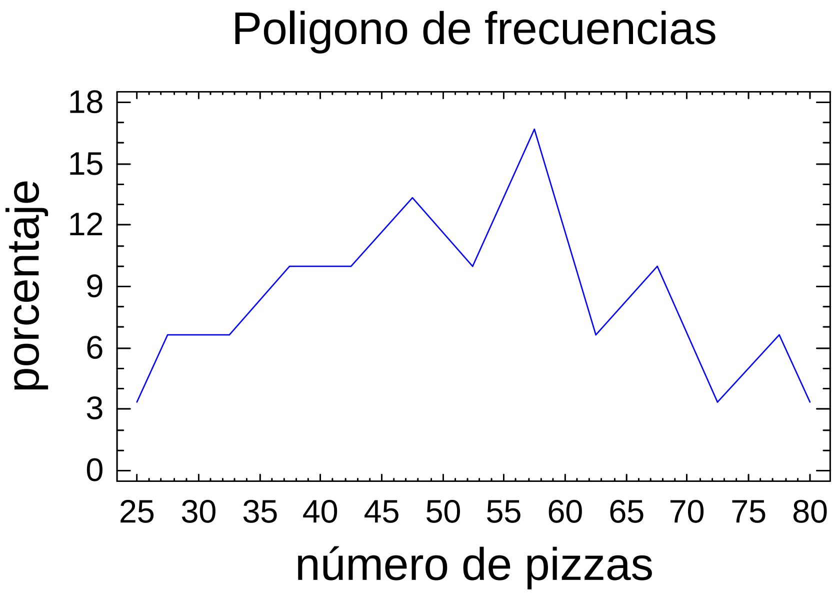
   $Q_1 = M_e = 40 + \frac{\frac{30}{4} - 7}{3} \times 5 = 40.833$

   $Q_2 = M_e = 51.667$

   $Q_3 = 62.5 + \frac{3 \times \frac{30}{4} - 22}{3} \times 5 = 63.333$

5. Utilizando la marca de clase y la frecuencia absoluta obtenemos el siguiente histograma y polígono de frecuencias.


Histograma de frecuencias
porcentaje
0
3
6
9
12
15
18
25 30 35 40 45 50 55 60 65 70 75 80
número de pizzas


Poligono de frecuencias
porcentaje
0
3
6
9
12
15
18
25 30 35 40 45 50 55 60 65 70 75 80
número de pizzas

Ejercicio 2 *Los alumnos de un cierto grupo han obtenido las siguientes calificaciones en un test:*

7, 5, 5, 0, 3, 9, 5, 3, 7, 9, 7, 5, 5, 3 ,3, 5, 3, 5, 5, 3.

1. *Hallar las tablas de frecuencias y de frecuencias acumuladas.*

2. *Diagrama de barras y poligono de frecuencia simple y el histograma de frecuencia acumulada.*

3. *Calcular la media, la mediana y la moda*

4. *Calcular el recorrido, la desviación media, la varianza, la desviación típica, la cuasivarianza o varianza muestral, la cuasidesviación y el coeficiente de variación.*

5. *Calcular el primer y tercer cuartil*

1.

x_i	$\mathbf{n}_i$	f_i	N_i	F_i
0	1	0.05	1	0.05
3	6	0.30	7	0.35
5	8	0.40	15	0.75
7	3	0.15	18	0.9
9	2	0.10	20	1
$Totales$	$N = 20$	1		

siendo:

$$f_i = \frac{n_i}{N}, \quad N_i = \sum_{j=1}^{i} n_j \quad F_i = \frac{N_i}{N} = \sum_{j=1}^{i} f_j$$

2. En los siguientes gráficos se ha añadido los valores que no aparecen,

que por lo tanto tendrán una frecuencia nula

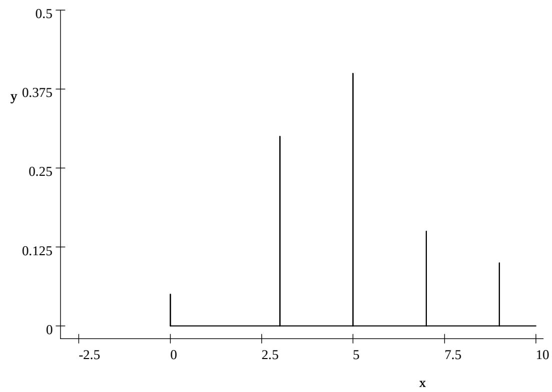
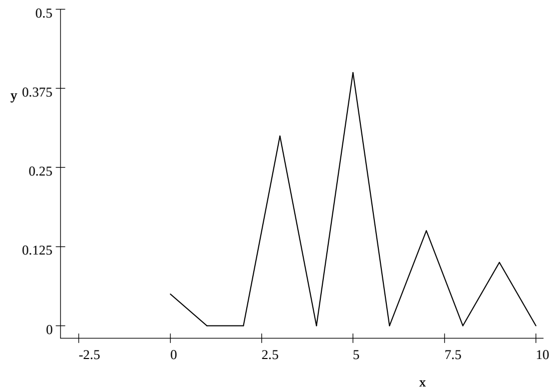


Diagrama de barras



Poligono de frecuencias

3. Resumimos los cálculos necesarios en la siguiente tabla:

x_i	$\mathbf{n}_i$	n_ix_i	$\lvert x_i - \overline{x}\rvert$	$\mathbf{n}_i\,\lvert x_i - \overline{x}\rvert$	$\mathbf{n}_i\,(x_i - \overline{x})^2$
0	1	0	-4.85	4.85	23.5225
3	6	18	-1.85	11.101	20.535
5	8	40	0.15	1.20	0.18
7	3	21	2.15	6.45	13.8675
9	2	18	4.15	8.3	34.445
Totales	20	97		31.9	92.55

media $= \frac{1}{N}\sum_{i=1}^{k} n_ix_i = \frac{97}{20} = 4.85$

Ordenando los elementos de la muestra de menor a mayor los elementos que ocupan los lugares centrales (décimo y undécimo) toman el valor 5, así que este es el valor de la mediana.

El elemento que tiene mayor frecuencia es 5, así que la moda también vale 5.

4. Recorrido $= 9 - 0 = 9$

   desviación media $= \frac{1}{N}\sum_{i=1}^{k} n_i\,|x_i - \overline{x}| = \frac{31.9}{20} = 1.595$

   varianza $= S^2 = \frac{1}{N}\sum_{i=1}^{k} n_i\,(x_i - \overline{x})^2 = \frac{92.55}{20} = 4.6275$

   desviación típica $= S = \sqrt{S^2} = \sqrt{4.6275} = 2.1512$

   Cuasivarianza $= s^2 = \frac{1}{N-1}\sum_{i=1}^{k} n_i\,(x_i - \overline{x})^2 = \frac{92.55}{19} = 4.8711$

   Cuasidesviación $= s = \sqrt{s^2} = \sqrt{4.8711} = 2.2071$

   Coeficiente de variación $= \frac{\textit{desviación típica}}{\textit{media}} = \frac{2.1512}{4.85} = 0.44355$

5. El primer cuartil está entre los términos que ocupan los lugares quinto y sexto. Ambos toman el valor 3, así que el valor del primer cuartil es 3.

   El tercer cuartil está entre el término 15º que es 5 y el el término 16º que vale 7. Si se toma el valor promedio entre ambos resulta 6 para el valor del tercer cuartil.

Ejercicio 3 *La siguiente tabla muestra los tiempos en segundos empleados en establecer una conexión a Internet en 75 ocasiones. Los datos se han agrupados mediante la siguiente tabla de frecuencias.*

Tiempos en seg.	$\mathbf{n}_i$
1,30 – 1.35	3
1.35 – 1.40	6
1,40 – 1.45	6
1.45 – 1.50	14
1,50 – 1.55	12
1.55 – 1.60	13
1,60 – 1.65	15
1.65 – 1.70	2
1,70 – 1.75	2
1.75 – 1.80	2
Total	$N = 75$

1. *Construir el histograma de frecuencias. Indicar el intervalo modal.*

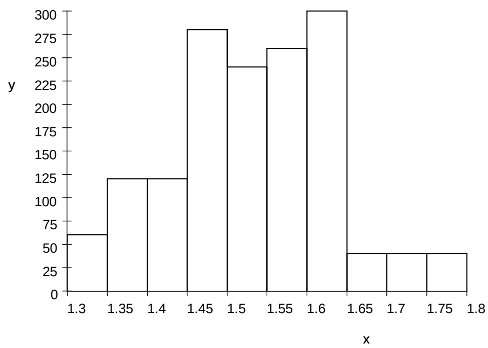
2. *Calcular la media y la mediana.*

3. *Calcular la varianza y el recorrido intercuartílico.*

1. La amplitud de cada clase es 0.05. Para calcular la altura de cada barra del histograma usamos la expresión:

$$h_i = \frac{n_i}{0.05}$$

Intervalos de clase	*Marcas de clase* x_i	$\mathbf{n}_i$	$h_i = \frac{n_i}{0.05}$
$1,30 \; - 1.35$	1.325	3	60
$1.35 \; - 1.40$	1.375	6	120
$1,40 \; - 1.45$	1.425	6	120
$1.45 \; - 1.50$	1.475	14	280
$1,50 \; - 1.55$	1.525	12	240
$1.55 \; - 1.60$	1.575	13	260
$1,60 \; - 1.65$	1.625	15	300
$1.65 \; - 1.70$	1.675	2	40
$1,70 \; - 1.75$	1.725	2	40
$1.75 \; - 1.80$	1.775	2	40
Total		$N = 75$	



Histograma de frecuencias

El intervalo modal es el que alcanza mayor altura. En este caso 1.60 - 1.65

2. Los cálculos necesarios se resumen en la tabla siguiente:

Intervalos de clase	*Marcas de clase* x_i	$\mathbf{n}_i$	$n_i x_i$	$x_i - \overline{x}$	$\mathbf{n}_i (x_i - \overline{x})^2$
$1,30 - 1.35$	1.325	3	3.975	−0.2074	0.130
$1.35 - 1.40$	1.375	6	8.25	−0.1574	0.149
$1,40 - 1.45$	1.425	6	8.55	−.01074	6.92×10^{-2}
$1.45 - 1.50$	1.475	14	20.65	−0.0574	4.61×10^{-2}
$1,50 - 1.55$	1.525	12	18.3	−0.0074	6.57×10^{-4}
$1.55 - 1.60$	1.575	13	20.475	0.0426	2.36×10^{-2}
$1,60 - 1.65$	1.625	15	24.375	0.0926	0.12862
$1.65 - 1.70$	1.675	2	3.35	0.1426	0.04067
$1,70 - 1.75$	1.725	2	3.45	0.1926	0.07419
$1.75 - 1.80$	1.775	2	3.55	0.2426	.0.11771
Total		75	114.93		0.75487

$media = \frac{1}{N}\sum_{i=1}^{k} n_i x_i = \frac{114.93}{75} = 1.5324$

La mediana ha de dejar delante el 50% de los datos. Por lo tanto el intervalo mediano resulta ser $1.50 - 1.55$

Aplicando la fórmula de interpolación lineal entre los extremos de este intervalo resulta:

$Mediana = 1.50 + \frac{\frac{75}{2} - 29}{12} 0.05 = 1.5354$

3. $Varianza = \frac{1}{N}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \frac{0.75487}{75} = 1.0065 \times 10^{-2}$

   El recorrido intercuartílico es la diferencia entre los valores del tercer y primer cuartil

   Cálculo del primer cuartil: El primer cuartil ha de dejar por delante el 25% de los datos. en este caso el lugar aproximado es en el dato $\frac{75}{4} = 18.75$ que corresponde al intervalo de clase $1.45 - 1.50$

   $Cuartil\ 1^o = 1.45 + \frac{\frac{75}{4} - 15}{14} 0.05 = 1.4634$

   En el caso del tercer cuartil $3 \times \frac{75}{4} = 56.25$

   $Cuartil\ 3^o = 1.60 + \frac{3 \times \frac{75}{4} - 54}{15} 0.05 = 1.6075$

   Por lo tanto el recorrido intercuartílico es $= 1.6075 - 1.4634 = 0.1441$

Ejercicio 4 *La distribución correspondiente al peso en kilos de 100 mujeres de 20 años viene resumida en la siguiente tabla de frecuencias:*

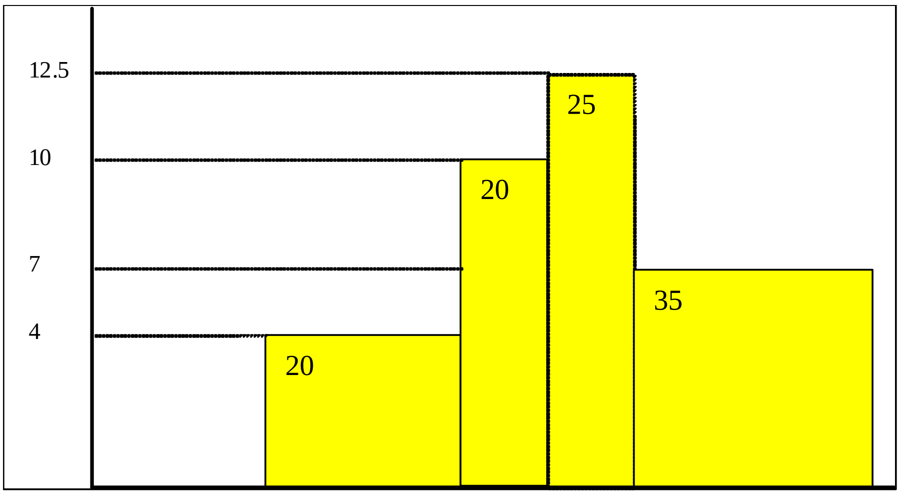
Intervalos de la variable	*fr. absolutas)*
[60, 65)	*20*
[65, 67)	*20*
[67, 69)	*25*
[69, 74)	*35*

1. *Forma la tabla de frecuencias de la distribución detallando, aparte de las frecuencias absolutas (n_i) el intervalo de clase, $[x_i, x_{i+1})$, la marca de clase (a_i), la amplitud del intervalo, $(x_{i+1} - x_i)$, la altura (h_i) del histograma de frecuencia para cada clase, así como las frecuencias relativas (f_i), absolutas acumuladas (N_i) y relativas acumuladas(F_i).*
2. *Representa los datos en un histograma*
3. *Estima cúantas mujeres pesan menos de 72 kilos*
4. *Determina la moda y la mediana.*
5. *Halla el rango intercuartílico, explicando su significado.*
6. *A partir de qué valor se encuentra el 25% de las mujeres con más peso?*

1.

$[x_i, x_{i+1})$	n_i	a_i	$x_{i+1} - x_i$	n_i	$h_i = \frac{n_i}{x_{i+1}-x_i}$	f_i	N_i	F_i
[60, 65)	20	62.5	5	20	4	0.2	20	0.2
[65, 67)	20	66	2	20	10	0.2	40	0.4
[67, 69)	25	68	2	25	12.5	0.25	65	0.65
[69, 74)	35	71.5	5	35	7	0.35	100	1

2.



3. Para responder esta cuestión suponemos que la población se distribuye "por igual dentro de cada intervalo". Por tanto, estarán por debajo de las 72 kilos todas las mujeres incluidas en los intervalos anteriores al [69, 74) $20 + 20 + 25 = 65$ y de este intervalo habrá que considerar la parte proporcional: En definitiva serán $65 + \frac{3}{5} \times 35 = 86$ mujeres.

4. El intervalo modal es el [67, 69). Si seleccionamos como moda la marca de clase de este intervalo la moda sería 68. Otra opción es emplear la expresión:

$$\begin{aligned} Mo &= L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i \\ \Delta_1 &= h_i - h_{i-1}, \quad \Delta_2 = h_i - h_{i+1}, \end{aligned}$$

Con ella el valor de la moda resulta:

$$Mo = 67 + \frac{0.125 - 0.1}{0.125 - 0.1 + 0.125 - 0.07} \times 2 = 67.625$$

El intervalo mediano es el [67, 69). Considerando que los valores se distribuyen por igual dentro de cada uno de los intervalos, el valor de la mediana se calcula como:

$$Me = L_i + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i = 67 + \frac{50 - 40}{25} \times 2 = 67.8$$

:

5. El rango intercuartílico es $Q3 - Q1$

$$Q_1 = L_i + \frac{\frac{n}{4} - N_{i-1}}{n_i} a_i Q_1 = 65 + \frac{25 - 20}{20} \times 2 = 65.5$$

$$Q_3 = L_i + \frac{\frac{3n}{4} - N_{i-1}}{n_i} a_i = 69 + \frac{75 - 65}{35} \times 5 = 70.429$$

$Q3 - Q1 = 70.429 - 65.5 = 4.929$

6. El valor viene dado por $Q_3 = 70.429$. Hay 25 mujeres que pesan más de 70.429 kilos.

Ejercicio 5 *Las calificaciones obtenidas por 40 alumnos/as de Bachillerato en las asignaturas de Matemáticas y las horas de estudio semanales que dedican a esta materia figuran en la siguiente tabla estadística bidimensional. En ella, la variable X hace referencia a la calificación lograda en Matemáticas e*

Y, al número de horas de estudio a la semana.

Y_j \ X_i	3	4	5	6	7	8	10	
2	*4*							*4*
5		*7*	*11*					*18*
6				*5*	*3*			*8*
7				*5*	*2*			*7*
9						*1*		*1*
10							*2*	*2*
Total	*4*	*7*	*11*	*10*	*5*	*1*	*2*	*40*

1. *Utilizar una tabla simple y otra de doble entrada para hallar la covariaza y el coeficiente de correlación lineal de Pearson*
2. *Analiza el grado de dependencia entre las calificaciones y las horas dedicadas al estudio.*
3. *En caso de que exista correlación ¿qué nota cabe esperar en Matemáticas un alumno que dedica 8 horas semanales al estudio?*
4. *¿Cúantas horas se estima que dedica al estudio un alumno que haya obtenido un 5 en Matemáticas?*

1. Comenzamos utilizando una tabla simple.

x_i	y_i	n_i	n_ix_i	n_iy_i	$n_ix_i^2$	$n_iy_i^2$	$n_ix_iy_i$
3	2	4	12	8	36	16	24
4	5	7	28	35	112	175	140
5	5	11	55	55	275	275	275
6	6	5	30	30	180	180	180
6	7	5	30	35	180	245	210
7	6	3	21	18	147	108	126
7	7	2	14	14	98	98	98
8	9	1	8	9	64	81	72
10	10	2	20	20	200	200	200
Sumas Totales		40	218	224	1292	1378	1325

$\overline{x} = \frac{218}{40} = 5.45$; $\overline{y} = \frac{224}{40} = 5.6$; $S_x = \sqrt{\frac{1292}{40} - 5.45^2} = 1.612$; $S_y = \sqrt{\frac{1378}{40} - 5.6^2} = 1.758$

$S_{xy} = \frac{1325}{40} - 5.45 \times 5.6 = 2.605$

El coeficiente de correlación es:

$r = \frac{S_{xy}}{S_x S_y} = \frac{2.605}{1.612 \times 1.758} = 0.919$

Presentamos ahora los cálculo en una tabla de doble entrada. Los cálculos para la marginal de X están en horizontal y los de la marginal de Y en vertical.

X_i / Y_j	3	4	5	6	7	8	10	$n_{\bullet j}$	$n_{\bullet j} y_j$	$n_{\bullet j} y_j^2$	$\sum_{j=1}^{7} n_{ij} x_i y_j$
2	4							4	8	16	24
5		7	11					18	90	450	415
6				5	3			8	48	288	306
7				5	2			7	49	343	308
9						1		1	9	81	72
10							2	2	20	200	200
$n_{i\bullet}$	4	7	11	10	5	1	2	40	224	1378	1325
$n_{i\bullet} y_i$	12	28	55	60	35	8	20	218			
$n_{i\bullet} y_i^2$	36	112	275	360	245	64	200	1292			
$\sum_{i=1}^{6} n_{ij} x_i y_j$	24	140	275	306	308	72	200	1325			

2 Como el coeficiente de correlación es cercano a uno se interpreta que hay un cierto tipo de relación estadística entre las variables. Como la covarianza es positiva las dos variables aumentan o disminuyen simultaneamente.

3 Empleamos la regresión de x sobre y (la calificación en función de las horas de estudio):

$x - \overline{x} = \frac{S_{xy}}{S_y^2}(y - \overline{y})$; $x - 5.45 = \frac{2.605}{1.758^2}(y - 5.6)$; $x = 5.45 + \frac{2.605}{1.758^2}(y - 5.6) = 5.45 + 0.842\,89(y - 5.6) = 0.842\,89y + 0.729\,83$

Para y=8 horas, la calificación esperada sería:

$x = 0.842\,89 \times 8 + 0.729\,83 = 7.\,473$

4 Ahora la recta que hay que calcular es la recta de regresión de "y" sobre "x":

$x = 5.6 + \frac{2.605}{1.612^2}(y - 5.45) = 1.\,002\,5y + 0.136\,46$

Sustituyendo la nota 5 en y, se obtiene:

$x = 1.\,002\,5y + 0.136\,460 = 1.\,002\,5 \times 5 + 0.136\,46 = 5.\,149\,0$

Un alumno que haya obtenido un 5 en Matemáticas, se estima que debe estudiar de 5.149.horas aproximadamente.

Ejercicio 6 *Ajustar una recta a los puntos dados en la siguiente tabla por medio de sus coordenadas x e y. Estudiar la calidad del ajuste*

x	y
1.444	0.565
1.898	0.132
1.171	2.584
7.006	-1.84
8.337	-4.12
9.323	-7.2
19.13	-119
11.63	-16.1
12.13	-17.3
17.55	-82.6
28.11	-393
32.29	-603
29.44	-449
34.88	-764
33.72	-686
35.35	-791
40.31	-1187

x	y	n	x^2	y^2	xy
1.444	0.565	1	2.085136	0.319225	0.81586
1.898	0.132	1	3.602404	0.017424	0.250536
1.171	2.584	1	1.371241	6.677056	3.025864
7.006	-1.84	1	49.084036	3.3856	-12.89104
8.337	-4.12	1	69.505569	16.9744	-34.34844
9.323	-7.2	1	86.918329	51.84	-67.1256
9.13	-119	1	365.9569	14161	-2276.47
11.63	-16.1	1	135.2569	259.21	-187.243
12.13	-17.3	1	147.1369	299.29	-209.849
17.55	-82.6	1	308.0025	6822.76	-1449.63
28.11	-393	1	790.1721	154449	-11047.23
2.29	-603	1	1042.6441	363609	-19470.87
29.44	-449	1	866.7136	201601	-13218.56
34.88	-764	1	1216.6144	583696	-26648.32
33.72	-686	1	1137.0384	470596	-23131.92
35.35	-791	1	1249.6225	625681	-27961.85
40.31	-1187	1	1624.8961	1408969	-47847.97
Totales: 323.719	-5117.879	17	9096.621	3830222.5	-173560.19

$N = 17$
$Media\ X = \frac{323.719}{17} = 19.042 \qquad Var(X) = \frac{9096.621}{17} - 19.042^2 = 172.50$
$S_x = \sqrt{172.50} = 13.134$
$MediaY = \frac{-5117.879}{17} = -301.05 \qquad Var(Y) = \frac{3830222.5}{17} - (-301.05)^2 = 1.3468 \times 10^5$
$S_y = \sqrt{1.3468 \times 10^5} = 366.99$
$Cov(X,Y) = \frac{-173560.19}{17} - 19.042\,(-301.05) = -4476.8$

La pendiente de recta de regresión es: $\frac{S_{XY}}{S_X^2} = \frac{-4476.8}{172.50} = -25.952$.

El término independiente lo deduciomos aplicando la propiedad de que la recta de regresión pasa por el centro de gravedad de los puntos:

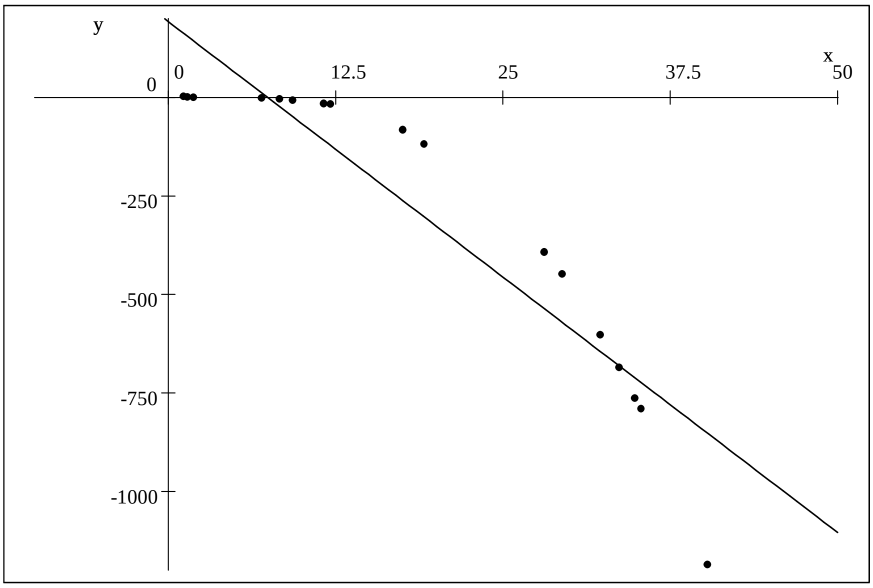
$-301.05 = -25.952 \times 19.042 + b$: Por tanto $b = 193.13$ y la recta de regresión resulta:

$$y = -25.952x + 193.13$$

El coeficiente de correlación lineal vale $\rho = \frac{S_{XY}}{S_x S_y} = \frac{-4476.8}{13.134 \times 366.99} = -0.92879$. El coeficiente de determinación es el cuadrado del de correlación: $\rho^2 = \left(\frac{S_{XY}}{S_x S_y}\right)^2 = (-0.92879)^2 = 0.86265$.

El coeficiente de determinación, así como el de correlación son próximos a 1, así que la recta de regresión se ajuste bastante bien a los puntos.

Se muestra a continuación la representación gráfica del ajuste



Ejercicio 7 *Dada la siguiente tabla de valores correspondiente a una variable*

estadística bidimensional (X,Y)

X	1	1	2	2	4	4	5	5
Y	1	5	2	4	2	4	1	5

calcular la recta de regresión de Y con respecto a X y el coeficiente de correlación. ¿Son X e Y incorreladas? ¿ Son X e Y independientes?

$$\overline{x} = \frac{1+1+2+2+4+4+5+5}{8} = 3.0, \qquad \overline{y} = \frac{1+5+2+4+2+4+1+5}{8} = 3.0$$
$$S_x^2 = \frac{(1-3)^2+(1-3)^2+(2-3)^2+(2-3)^2+(4-3)^2+(4-3)^2+(5-3)^2+(5-3)^2}{8} = 2.5$$
$$S_x^2 = \frac{(1-3)^2+(5-3)^2+(2-3)^2+(4-3)^2+(2-3)^2+(4-3)^2+(1-3)^2+(5-3)^2}{8} = 2.5$$
$$S_{xy} = \frac{(1-3)(1-3)+(1-3)(5-3)+(2-3)(2-3)+\ldots \quad \ldots+(4-3)(4-3)+(5-3)(1-3)+(5-3)(5-3)}{8} = 0$$

La recta de Regresión de Y con respecto a X tiene la expresión:

$$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

En este caso sería:

$$y - 3 = \frac{0}{2.5}(x-3), \text{ es decir la recta } y = 3$$

El coeficiente de correlación es

$$r = \frac{S_{xy}}{S_x S_y} = \frac{0}{2.5} = 0$$

por tanto las variables son incorreladas. Las variables X *e* Y no son independientes, ya que las distribuciones de $Y/X = x_i$ no coinciden. Analogamente sucede con las distribuciones de $X/Y = y_j$

Ejercicio 8 *La tabla siguiente muestra los mejores tiempos mundiales en Juegos Olimpicos hasta 1976 en carrera masculina para distintas distancias. La variable y registra el tiempo en segundos y la variable x la distancia recorrida en metros.*

y	*9.9*	*19.8*	*44.26*	*103.5*	*214.9*	*806.4*	*1658.4*	*7795*
x	*100*	*200*	*400*	*800*	*1500*	*5000*	*10000*	*42196*

1. *Calcular la recta de regresión de y sobre x*

2. *Calcular la varianza residual y el coeficiente de correlación. Indicar si el ajuste lineal es adecuada, usando este último coeficiente.*

1. $\overline{x}=\frac{100+200+400+800+1500+5000+10000+42196}{8}=7524.5$

   $\overline{y}=\frac{9.9+19.8+44.26+103.5+214.9+806.4+1658.4+7795}{8}=1331.5$

   $S_x^2=\frac{(100-7524.5)^2+(200-7524.5)^2+...+(42196-7524.5)^2}{8}=1.82\times 10^8$

   $S_y^2=\frac{(9.9-1331.5)^2+(19.8-1331.5)^2+... \quad ...+(7795-1331.5)^2}{8}=6.2548\times 10^6$

   $S_{xy}=\frac{(100-7524.5)(9.9-1331.5)+. \quad ...+(42196-7524.5)(7795-1331.5)}{8}$

   $=3.3726\times 10^7$

   La pendiente es $\frac{3.3726\times 10^7}{1.8196\times 10^8}=.18535$

   El término constante es

   $\overline{y}-0.185\overline{x}=1331.5-0.18535\times 7524.5=-63.166$

   Por tanto la recta de regresión es

   $y=0.18535\,x-63.166$

2. La varianza residual es

   $S_e^2=\frac{\sum_1^n(e_i-\overline{e})^2}{n}=\frac{\sum_1^n {e_i}^2}{n}=\frac{\sum_1^n(y_i-(0.18535\,x-63.166))^2}{n}$. Realizamos las operaciones en forma matricial

$$X=\begin{pmatrix}100\\200\\400\\800\\1500\\5000\\10000\\42196\end{pmatrix},\quad Y=\begin{pmatrix}9.9\\19.8\\44.26\\103.5\\214.9\\806.4\\1658.4\\7795\end{pmatrix}$$

$$Y^*=0.18535\begin{pmatrix}100\\200\\400\\800\\1500\\5000\\10000\\42196\end{pmatrix}-\begin{pmatrix}63.166\\63.166\\63.166\\63.166\\63.166\\63.166\\63.166\\63.166\end{pmatrix}=\begin{pmatrix}-44.631\\-26.096\\10.974\\85.114\\214.86\\863.58\\1790.3\\7757.9\end{pmatrix}$$

$$e = Y - Y^* = \begin{pmatrix} 9.9 \\ 19.8 \\ 44.26 \\ 103.5 \\ 214.9 \\ 806.4 \\ 1658.4 \\ 7795 \end{pmatrix} - \begin{pmatrix} -44.631 \\ -26.096 \\ 10.974 \\ 85.114 \\ 214.86 \\ 863.58 \\ 1790.3 \\ 7757.9 \end{pmatrix} = \begin{pmatrix} 54.531 \\ 45.896 \\ 33.286 \\ 18.386 \\ .04 \\ -57.18 \\ -131.9 \\ 37.1 \end{pmatrix}$$

$$\sum_1^n {e_i}^2 = \begin{pmatrix} 54.531 & 45.896 & 33.286 & 18.386 & .04 & -57.18 & -131.9 & 37.1 \end{pmatrix} \begin{pmatrix} 54.531 \\ 45.896 \\ 33.286 \\ 18.386 \\ .04 \\ -57.18 \\ -131.9 \\ 37.1 \end{pmatrix} = 28570.0$$

Por tanto la varianza residual es
$\frac{\sum_1^8 {e_i}^2}{8} = \frac{28570}{8} = 3571.3$

3. El coeficiente de correlación es $\frac{S_{xy}}{s_x s_y} = \frac{3.3726\times10^7}{\sqrt{6.2548\times10^6}\sqrt{1.8196\times10^8}} = .9997$

   Por lo tanto el coeficiente de correlación parece indicar un buen ajuste lineal.

Ejercicio 9 *Las rectas de regresión de una distribución son: la de Y sobre X, $y = 0.5x + 2$ y la de X sobre Y, $x = 1.8y + 5$. hallar:*

1. *El coeficiente de correlación*
2. *El centro de gravedad de los puntos*

1. El térrmino independiente de la primera recta es $\frac{S_{XY}}{S_X^2} = 0.5$ y el de la segunda $\frac{S_{XY}}{S_Y^2} = 1.8$

   Hallamos el coeficiente de determinación a partir de las pendientes de las rectas de regresión. Multiplicando ambos términos obtenemos el cuadrado del coeficiente de correlación, llamado coeficiente de determinación:

   $\frac{S_{XY}}{S_X^2}\frac{S_{XY}}{S_Y^2} = \left(\frac{S_{XY}}{S_x S_Y}\right)^2 = 0.5 \times 1.8 = 0.9$. Por tanto el coeficiente de correlación es $\sqrt{0.9} = 0.94868$.que al encontrarse próximo a 1 indica que ambas rectas son un buen ajuste para la nube de puntos.

2. Las dos rectas de regresión deben pasar por el centro de gravedad , bastará resolver el sistema formado por las dos ecuaciones:

$\begin{cases} y = 0.5x + 2 \\ x = 1.8y + 5 \end{cases}$; $x = 1.8(0.5x + 2) + 5 = 0.9x + 8.6; 0.1x = 8.6; x = 86$, $y = 43 + 2 = 45$

El centro de gravedad es el punto (86,45)

Ejercicio 10 *Una empresa inmobiliaria ofrece apartamentos en régimen de alquiler, cuyos precios mensuales y número de ellos para cada intervalo de precio son:*

Precio de alquiler	*Nº de apartamentos*
De 700 a 1000	*21*
1000 a 1100	*27*
1100 a 1300	*34*
1300 a 1500	*14*
1500 a 1800	*8*

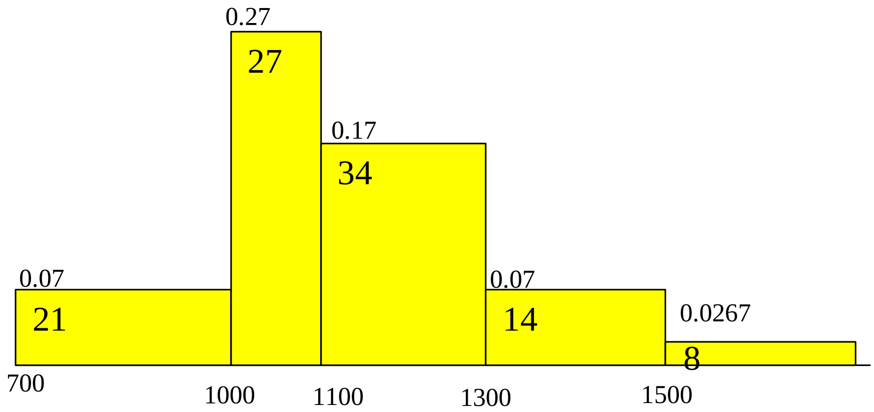
1. *Completar la tabla de frecuencias y realizar la representación gráfica más adecuada.*
2. *Obtener los coeficientes de centralización y dispersión.*
3. *Si una persona quiere gastar en alquiler entre 1250 y 1350 euros al mes, aproximadamente, ¿a qué porcentaje del total de apartamentos tiene opción?*

1.

Precio de alquiler	x	Nº de apartamentos	h	n*x
De 700 a 1000	850	21	0.07	17850
1000 a 1100	1050	27	0.27	28350
1100 a 1300	1200	34	0.17	40800
1300 a 1500	1400	14	0.07	19600
1500 a 1800	650	8	0.026666667	13200
SUMAS		104		119800

|x-media|	n|x-media|	n(x-media)^2
350	7350	2572500
150	4050	607500
0	0	0
200	2800	560000
450	3600	1620000
	17800	5360000

La representación gráfica más adecuada es el histograma. Nótese que la frecuencia de cada intervalo es proporcional al area del rectángulo que lo tiene como base.



2. **Precio medio** $= \quad \frac{119800}{104} = 1151.9$

Moda: El intervalo modal es el segundo. Su marca (1050) de clase podría darse como moda.

Otra forma es mediante la fórmula:

$$Mo = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} a_i = 1000 + \frac{0.17}{0.17 + 0.07} 100 = 1070.8$$

Mediana: El intervalo mediano es el que contiene los valores de orden 54 y 55, que corresponde a 1100-1300. Su marca de clase, 1200, puede darse como mediana. Otra forma es aplicar la fórmula:

$$Me = L_i + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i = 1100 + \frac{52 - 48}{34} 200 = 1123.5$$

Desviación media

$$dm = \frac{\sum_{i=1}^{5} n_i \left| x_i - \bar{x} \right|}{n} = \frac{17800}{104} = 171.15$$

Desviación típica

$$\sigma = \sqrt{\frac{\sum_{i=1}^{5} n_i \left(x_i - \bar{x}\right)\hat{}^2}{n}} = \sqrt{\frac{5360000}{104}} = 227.02$$

Coeficiente de variación: $\frac{227.02}{1151.9} = 0.19708$

3. $Porcentaje = \frac{1300-1250}{1300-1100} \times 34 + \frac{1350-1300}{1500-1300} \times 14 = 12.0$ elementos corresponden proporcionalmente a dicho intervalo. En porcentaje sería:

$\frac{12}{104} \times 100 = 11.55\%$

Ejercicio 11 *Dada la siguiente tabla de doble entrada, que recoge las frecuencias observadas para la variable bidimensional (X,Y) (X en horizontal, Y en vertical)*

Y\ X	*5*	*10*	*15*	*20*
10	*30*	*2*	*0*	*0*
20	*0*	*23*	*1*	*0*
30	*0*	*2*	*8*	*0*
40	*0*	*0*	*2*	*4*

Se pide:

1. *Distribución marginal de las variables X e Y.*
2. *Varianza de Y y covarianza de (X,Y).*
3. *Coeficiente de determinación.*

1. La marginal de ambas variables viene dada en las tablas:

	ni
10	32
20	24
30	10
40	6
Total	72

	5	10	15	20		ni
nj	30	27	11	4		Total = 72

2. $MediaX = \frac{1340}{72} = 18.611$ $\quad Var(X) = \frac{31400}{72} - 18.611^2 = 89.742$ $Desv(X) = \sqrt{89.742} = 9.4732$

$MediaY = \frac{665}{72} = 9.2361$ $\quad Var(Y) = \frac{7525}{72} - 9.236^2 = 19.21$ $\quad Desv(Y) = \sqrt{19.21} = 4.383$

$\sum x_i y_j = 15200$ $\qquad \frac{\sum x_i y_j}{n} = \frac{15200}{72} = 211.11$

$Cov(X;Y) = \frac{\sum x_i y_j}{n} - \overline{\overline{x}y} = 211.11 - 18.611 \times 9.2361 = 39.217$

3. $Coef.Correl = \frac{Cov(X;Y)}{S \times S_y} = \frac{39.217}{9.4732 \times 4.383} = 0.94451$

   $Coef.Determ = 0.94451^2 = 0.892$

Ejercicio 12 *Al realizar un estudio para comprobar la relación entre el tiempo en días tardado en diferentes empresas en el desarrollo de sus aplicaciones informáticas y en su posterior implantación se obtuvieron los siguientes resultados:*

Desarrollo (X)	*75*	*80*	*93*	*65*	*87*	*71*
Implantación (Y)	*82*	*78*	*86*	*73*	*91*	*80*

1. *Hallar la recta de regresión de Y respecto de X.*
2. *Realizar un ajuste del tipo $Y = ab^X$*
3. *¿Qué ajuste te parece más conveniente?*

1.

X		Y	ni	X^2	Y^2	XY
75		82	1	5625	6724	6150
80		78	1	6400	6084	6240
93		86	1	8649	7396	7998
65		73	1	4225	5329	4745
87		91	1	7569	8281	7917
71		80	1	5041	6400	5680
Totales	471	490	6	37509	40214	38730

$N = 6$

$MediaX = \frac{471}{6} = 78.5 \qquad Var(X) = \frac{37509}{6} - 78.5^2 = 89.25$

$desv(x) = \sqrt{89.25} = 9.4472$

$MediaY = \frac{490}{6} = 81.667 \qquad Var(Y) = \frac{40214}{6} - 81.667^2 = 32.834 = 88889$

$desv(y) = \sqrt{32.834} = 5.7301$

$cov(X,Y) = \frac{38730}{6} - 78.5 \times 81.667 = 44.141$

$a = \frac{Cov(X,Y)}{S_x^2} = \frac{44.141}{89.25} = 0.49458 = 81.667, \ b = 81.667 - 0.49458 \times 78.5 = 42.842$

La recta de regresión es $Y = \quad 42.842 + 0.49458X$

El coeficiente de correlación lineal vale

$\frac{Cov(X,Y)}{S_x S_y} = \frac{44.141}{9.4472 \times 5.7301} = 0.815\,41$

El coeficiente de determinación vale $0.815\,41^2 = 0.664\,89$

2. $Y = ab^X$

Tomando log en base 10, obtenemos

$logY = loga + X \log b.$

Ahora ajustamos una recta a los valores de X y $logY$

X	logY	ni	X^2	Y^2	XY
75	1.913813852	1	5625	3.662683462	143.5360389
80	1.892094603	1	6400	3.580021986	151.3675682
93	1.934498451	1	8649	3.742284258	179.908356
65	1.86332286	1	4225	3.471972081	121.1159859
87	1.959041392	1	7569	3.837843177	170.4366011
71	1.903089987	1	5041	3.621751499	135.1193891
Totales: 471	11.46586115	6	37509	21.91655646	901.4839392

Repitiendo los cálculos para la recta de regresión se obtiene:

$N = 6$

$MediaX = 78.5 \qquad Var(X) = 89.25 \qquad desv(x) = 9.447221814$

$MediaY = 1.910976858 \qquad Var(Y) = 0.00092686 \qquad desv(y) = 0..030444373$

$Cov(X,Y) = 0.235639881$

$lny = lna + x \times lnb$

La recta de regresión es $\quad Y = 1.703719371 + 0.002640223X$

De donde se deduce que

$a = 10^{1.703719371} = 50.\,550$

$b = 10^{0.002640223} = 1.\,006\,1$

Por tanto el ajuste es $Y = 50.550 \times 1.0061^X$

3. Hallamos el coeficiente de determinación del ajuste para compararlo con el ajuste lineal del primer apartado. Los cálculos necesarios están

dados en la tabla:

X	Y	$Y^* = ab^x$	errores al cuadrado
75	82	79.75098909	5.058050068
80	78	82.21237447	17.74409869
93	86	88.97337965	8.840986554
65	73	75.04708889	4.19057294
87	91	85.78646081	27.18099088
71	80	77.8350461	4.687025388
471	490	489.605339	67.70172452

coef determ $= 1 - \frac{var(error)}{var(y)} = 1 - \frac{\frac{67.70172452}{6}}{32.834} = 1 - \frac{11.28362075}{32.834} = 0.656\,34$

La recta es ligeramente mejor, ya que su coeficiente de determinación era 0.664 89.

T. 2

Cálculo de probabilidades

2.1 Repaso de combinatoria

Ejercicio 13 *¿ De cuantas formas pueden sacarse tres cartas de una baraja de 40 cartas?*

1. *Teniendo en cuenta el orden*
2. *Sin tener en cuenta el orden*

1. $V_{40,3} = 40 \times 39 * \times 38 = 59280$
2. $C_{40,3} = \frac{V_{40,3}}{P_3} = \frac{59280}{3.2.1} = 9880$

Ejercicio 14 *Si sacamos tres cartas de una baraja de 40 cartas ¿ De cuantas formas pueden sacarse una pareja?*

1. *Teniendo en cuenta el orden*
2. *Sin tener en cuenta el orden*

1. $10 \times (6) \times 36 \times 6 = 12960$
   (número posibles para las parejas)×(parejas diferentes del mismo número)×
   ×(maneras de sacar la carta diferente)×(cambios de orden)
2. $\frac{12960}{3!} = 2160$

Ejercicio 15 *Si sacamos tres cartas de una baraja de 40 cartas ¿ De cuantas formas pueden sacarse tres cartas del mismo número?*

1. *Teniendo en cuenta el orden*

2. Sin tener en cuenta el orden

$10 \times V_{4,3} = 10 \times (4 \times 3 * \times 2) = 240$
(números posibles para los trios)*(distintas formas de sacar cada trio con el mismo número)
$10 \times C_{4,3} = 10 \times \frac{V_{4,3}}{P_3} = 10 \times \frac{4*3*2}{3*2*1} = 40$

Ejercicio 16 *Si sacamos tres cartas de una baraja de 40 cartas ¿ De cuantas formas pueden sacarse tres cartas de distinto número?*

1. *Teniendo en cuenta el orden*

2. *Sin tener en cuenta el orden*

1. $40 \times 36 \times 32 = 46080$.

   Usando los resultados de los ejercicios anteriores 13, 14 y 15, podríamos hacerlo tambien restando del total las combinaciones que forman parejas y las que forman trio:

   $59280 - (12960 + 240) = 46080$.

2. $\frac{46080}{3!} = 7680$.

   De forma análoga a lo que se ha hecho en el primer apartado podemos calcularlo también como:

   $9880 - (2160 + 40) = 7680$

Ejercicio 17 *¿Cúantos números de 3 cifras pueden fomarse con los dígitos impares,1,3,5,7,9? ¿Y con los pares 0,2,4,6,8?*

En el primer caso son las variaciones con repetición de 5 elementos tomados en grupos de 3:
$V'_{5,3} = 5 \times 5 \times 5 = 5^3 = 125$.
En el segundo caso el prmer lugar no puede estar ocupado por el 0.
Por tanto serán:
$4 \times 5 \times 5 = 100$

Ejercicio 18 *¿De cuántas maneras pueden elegir 20 operarios de una fábrica una comisión formada por 3 de ellos, que los represente ante la Empresa?*

$\binom{20}{2} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140$.
Hay 1140 comisiones posibles.

Ejercicio 19 *¿Cuántas palabras diferentes pueden formarse con las letras de la palabra ESTADISTICA si las consonantes han de ocupar los lugares impares y las vocales los pares.*

Las consonantes son: S S T T D C y las vocales E A A I I. Los cambios de orden de las consonantes y de las vocales, al tener letras repetidas, pueden calcularse por medio de permutaciones con repetición.

Las posiciones posibles para las consonantes son :
$P_{6;2,2,1,1} = \frac{6!}{2!2!1!1!} = 180$
Las posiciones posibles para las vocales son.
$P_{5;1,2,2} = \frac{5!}{1!2!2!} = 30$
Luego se pueden formar $180 \times 30 = 5400$ palabras

2.2 Probabilidad

Ejercicio 20 *Si sacamos tres cartas de una baraja de 40 cartas ¿ Cual es la probabilidad de sacar una pareja?*

$\frac{12960}{59280} = 0.2186$

Ejercicio 21 *Si sacamos tres cartas de una baraja de 40 cartas ¿ Cual es la probabilidad de sacar un trio?*

$\frac{240}{59280} = 0.004$

Ejercicio 22 *Si sacamos tres cartas de una baraja de 40 cartas ¿ Cual es la probabilidad de sacar tres cartas de distinto número?*

$\frac{46080}{59280} = 0.7773$
Usando los resultados anteriores podríamos calcular la probabilidad del suceso contrario:1-(0.2186+0.004)=0.7774. La diferencia en la última cifra procede de los errores de redondeo.

Ejercicio 23 *¿ Son independientes los sucesos sacar una carta de oro y sacar un cuatro?*

Sea A ={sacar una carta de oro}, B = {sacar un cuatro}
P(A) = P(sacar una carta de oro)= $\frac{10}{40} = \frac{1}{4}$
P(A/B)=P{sacar una carta de oros/si es un cuatro}
$=\frac{P(sacar\, el cuatro\, de\, oro)}{P(sacar\, un cuatro)0} = \frac{\frac{1}{40}}{\frac{4}{40}} = \frac{1}{4}$
Como P(A/B) = P(A) los sucesos son independientes

Ejercicio 24 *En una fábrica hay 8000 obreros (80% hombres), 1500 administrativos (1000 mujeres y 500 hombres) y 500 personas que realizan labores de dirección (10% mujeres)? . Si elegimos una persona al azar,*

1. *¿ Cual es la probabilidad de que sea administrativo?*

2. *¿ Cual es la probabilidad de que sea una mujer?*

3. *¿ Cual es la probabilidad de que una mujer sea administrativo? ¿ Son independientes los sucesos ser administrativo y ser mujer?*

4. *¿ Son independientes los sucesos ser obrero y ser mujer?*

5. *Cual es la probabilidad de que una mujer sea directiva? ¿ Cual es la probabilidad de que hombre sea directivo?*

1. A={ser administrativo}, P(A) $= \frac{1500}{8000+1500+500} = \frac{1500}{10000} = 0.15$

2. M={ser mujer}, P(M)$= \frac{1600+1000+50}{10000} = \frac{2650}{10000} = 0.265$

   P(A/M)$= \frac{P(A \cap M)}{P(M)} = \frac{\frac{1000}{10000}}{\frac{2650}{1000}} = \frac{1000}{2650} = 0.377$

   $P(A/M) \neq P(A)$ luego los sucesos son dependientes. Hay más proporción de administrativos entre las mujeres (37.7%) que en el total de empleados (15%)

3. $P(O) = \frac{8000}{10000} = 0.80$, $P(O/M) = \frac{1600}{2650} = 0.60$. Los sucesos son dependientes, ya que P(O/M) $\neq$P(O)

4. $P(D/M) = \frac{\frac{50}{10000}}{\frac{2650}{10000}} = \frac{50}{2650} = 0.019$, $P(D/H) = \frac{450}{7350} = 0.061$

Ejercicio 25 *Tenemos tres urnas A,B,C . A tiene dos bolas blancas y una negra, B tiene dos bolas blancas y dos negras, C tiene 3 bolas blancas y una negra. Realizamos el experimento consistente en tirar un dado y sacar luego una bola de una urna. Si el resultado del dado es un número par elegimos la urna A y sacamos de ella una bola. Si el resultado es un 1 elegimos la urna B para sacar la bola. En los restantes casos sacamos la bola de la urna C.*

1. *¿ Cual es la probabilidad de que la bola extraida sea blanca?*

2. *Si se sabe que la bola resulto ser blanca, ¿ Cual es la probabilidad de que proceda de la primera urna?*

1. Usamos el teorema de la probabilidad total. Sea B_L el suceso sacar bola blanca y A,B,C el suceso correspondiente a elegir cada una de las respectivas urnas.

   $P(B_L)= P(A)*P(B_L/A)+ P(B)*P(B_L/B)+ P(C)*P(B_L/C) =$

   $\frac{1}{2} * \frac{2}{3} + \frac{1}{6} * \frac{1}{2} + \frac{2}{6} * \frac{3}{4} = \frac{2}{6} + \frac{1}{12} + \frac{6}{24} = 0.666$

2. Empleando el teorema de Bayes

   P(A/BB$_L$) $= \frac{P(A)*P(B_L/A))}{P(B_L)} =$

   $= \frac{P(A)*P(B_L/A))}{P(A)*P(B_L/A)+P(B)*P(B_L/B)+P(C)*P(B_L/C)} = \frac{\frac{1}{2}*\frac{2}{3}}{0.66} = 0.5$

Ejercicio 26 *En un sistema hay instalada una alarma. La probabilidad de que se produzca un peligro es 0.1. Si se produce, la probabilidad de que la alarma funcione es 0.95. La probabilidad de que la alarma funcione sin haber peligro es 0.03 . Hallar:*

1. *La probabilidad de que funcione la alarma*
2. *Probabilidad de que habiendo funcionado la alarma no haya habido peligro.*
3. *Probabilidad de que haya un peligro y, para colmo, la alarma no funcione.*
4. *Probabilidad de que no habiendo funcionado la alarma haya peligro.*

Sea P el suceso “hay peligro”, P^c “no hay peligro”, F “funciona la alarma”, F^c “no funciona la alarma”.

Los datos son: P(P)=0.1, p(F/P)=0.95, P(F/P^c)=0.03

1. $P(F) = P(P)P(F/P) + P(P^c)P(F/P^c) = 0.1 \times 0.95 + 0.9 \times 0.03 =$

   $= 0.095 + 0.027 = 0.122$

   $P(P^c/F) = \frac{P(P^C \cap F)}{P(F)} = \frac{P(F)*P(F/P^C)}{P(F)} = \frac{0.027}{0.122} = 0.22$

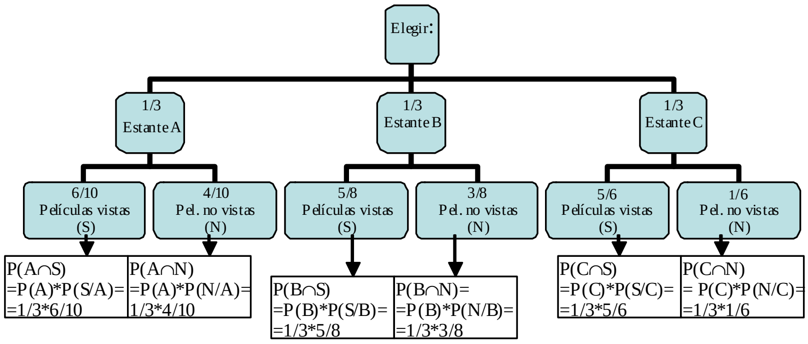
   $P(P \cap F^c) = P(P)P(F^c/P) = 0.1 \times (1 - P(F/P) = 0.1 \times 0.05 = 0.005$

   $P(P/F^c) = \frac{P(P \cap F^c)}{1-P(F)} = \frac{0.005}{0.88} = 0.006$

Ejercicio 27 *Tengo mis películas clasificada en tres estantes, A, B, y C. El estante A contiene 10 películas, de las cuales aún no he visto 4 de ellas; El B contiene 8 películas y no he visto 3 de ellas y el C contiene 6 películas de las cuales sólo me falta por ver una de ellas. Si selecciona al azar un estante y cojo una película.*

1. *¿Cuál es la probabilidad de que no la haya visto?*
2. *Si estoy viendo una película por primera vez, pero no recuerdo el estante del que procede, ¿cuál es la probabilidad de que provenga del estante A?*

1. Recojo toda la información en un grafo. Represento por N el suceso que consiste en elegir una pélicula que no se había visto previamente y S el que consiste en elegir una película ya vista.



La probabilidad de que haya elegido una película que no he visto es:

$$\begin{aligned} P(N) &= P(A)P(N/A) + P(B)P(N/B) + P(C)P(N/C) = \\ \frac{1}{3}\frac{4}{10} + \frac{1}{3}\frac{3}{8} + \frac{1}{3}\frac{1}{6} &= \frac{2}{15} + \frac{1}{8} + \frac{1}{18} = 0.313\,89 \end{aligned}$$

2. Para calcular la probabilidad de que la película que no he visto provenga del estante A, aplicamos el teorema de Bayes:

$$P(A/N) = \frac{P(A\cap N)}{P(N)} = \frac{P(A)P(N/A)}{P(N)} = \frac{\frac{2}{15}}{0.313\,89} = 0.424\,78$$

Ejercicio 28 *Sean A y B dos sucesos tales que $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ y $P(A\cap B) = \frac{1}{7}$. Calcular: $P(A/B)$, $P(A\cup B)$, $P(A'\cap B)$ y $P(A'\cap B')$*

1. $P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{7}}{\frac{1}{3}} = \frac{3}{7}$

$P(A\cup B) = P(A) + P(B) - P(A\cap B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{7} = \frac{29}{42}$
$P(A'\cap B) = P(B)P(A'/B) = P(B)\left[1 - P(A/B)\right] = \frac{1}{3}\left(1 - \frac{3}{7}\right) = \frac{4}{21}$
$P(A'\cap B') = P\left[(A\cup B)'\right] = 1 - P(A\cup B) = 1 - \frac{29}{42} = \frac{13}{42}$

Ejercicio 29 *En una encuesta sobre habitos de lectura se han sacado las siguientes conclusiones: El 55% de los ciudadanos no lee el periódico A. El 65% de los ciudadanos no lee el periódico B. El 10% de los ciudadanos no lee ni el periódico A ni el B. ¿Es esto posible?*

Denotemos por A el suceso que corresponde a la propiedad "leen el periódico A", y por B, "leen el periódico B". De los datos del enunciado, tenemos $fr(A') = 0,55$, $fr(B') = 0,65$ y $fr(A' \cap B') = 0,10$.. Si calculamos los que no leen al menos uno de los dos periódicos sería

$fr(A' \cup B') = fr(A') + fr(B') - fr(A' \cap B') = 0.55 + 0.65 - 0.10 = 1.1$

La frecuencia relativa no puede ser mayor que 1. Por tanto la encuesta contiene errores.

Ejercicio 30 *En una ciudad el 56% de sus habitantes adultos son hombres, de los cuales el 34% afirma ser fumador, mientras que de las mujeres el 40% se confiesa fumadora. Eligiendo una persona al azar,*

1. *Calcula la probabilidad de que sea fumadora.*
2. *Probabilidad de que si la persona seleccionada al azar es fumadora sea varón.*
3. *Probabilidad que sea varón y no fumador.*

Traduciendo los datos del problema a probabilidades de sucesos:

$P(V) = 0.56$, por tanto $P(M) = 1 - P(V) = 1 - 0.56 = 0.44$; $P(F/V) = 0.34$; $P(F/M) = 0.40$

1. Usando el teorema de probabilidad total se deduce:

   $P(F) = P(V)P(F/V) + P(M).P(F/HM) = 0.56 \times 0.34 + 0.44 \times 0.40 = 0.3664$

   El 36.64% de los habitantes de la ciudad fuman

2. Usamos ahora el teorema de Bayes:

   $P(V/F) = \frac{P(V \cap F)}{P(F)} = \frac{P(V)P(F/V))}{P(F)} = \frac{0.56 \times 0.34}{0.3664} = \frac{0.1904}{0.3664} = 0.51965.$

   El 51.965 de los fumadores son varones.

3. $P(V \cap F') = P(V)P(F'/V) = P(V)[1 - P(F/V)] = 0.56(1 - 0.34) = 0.3696$

El 36.96% de las personas son hombres y además no fuman.

Ejercicio 31 *En una universidad los estudiantes se matriculan en primer curso de ingeniería, ciencias, letras u otras especialidades. De entre ellos, acaban la carrera el 50% de ingeniería, el 70% de ciencias y el 80% de letras. El número de estos estudiantes que concluye la carrera en las otras especialidades representa el 10% del total de estudiantes matriculados en primero. Se sabe que el 20% se matriculan en ingeniería, el 15% en ciencias y el 50% en letras. Se supone que no se matriculan en más de una especialidad. Calcular:*

1. *Probabilidad de que un estudiante que se matricule en primero vaya a acabar la carrera y sea de ingeniería.*
2. *Qué porcentaje de alumnos matriculados en otras especialidades acaba la carrera?*
3. *Globalmente, ¿qué porcentaje de alumnos de primero se espera que vaya a acabar la carrera?*
4. *Probabilidad de que un estudiante que haya acabado la carrera sea de ingeniería.*

Formalizamos los datos del enunciado:
$P(A/I) = 0.50$, $P(A/C) = 0.70$, $P(A/L) = 0.80$, $P(A \cap O) = 0.10$.
$P(I) = 0.20$, $P(C) = 0.15$, $P(L) = 0.50$.
Por tanto la probabilidad de estudiar otras especialidades es $P(O) = 1 - 0.20 - 0.15 - 0.50 = 0.15$

1. $P(A \cap I) = P(I)P(A/I) = 0.20 \times 0.50 = 0.1$
2. $P(A/O) = \frac{P(A \cap O)}{P(O)} = \frac{0.1}{0.15} = \frac{2}{3} = 0.666\,67$. El 66.7% de estos alumnos acaba la carrera.
3. $P(A) = P(I)p(A/I) + P(C)p(A/C) + P(L)p(A/L) + P(O)p(A/O) = 0.20 \times 0.50 + 0.15 \times 0.70 + 0.50 \times 0.80 + 0.15 \times 0.666\,67. = 0.705$

   Acaban la carrera el 70.5% de los matriculados en primero
4. Hay que calcular
   $P(I/A) = \frac{P(A \cap I)}{P(A)} = \frac{0.1}{0.70} = 0.142\,86.$

   El 14.28% de los estudiantes que acaba la carrera son de Ingeniería.

Ejercicio 32 *Se ha comprobado que el 40% de las personas que toman ciertos productos farmaceútico sufren efectos secundarios. En un grupo de 15 personas que toman estos productos, ¿Cúal es la probabilidad de que exactamente 4 de ellas sufran efectos secundarios? ¿Cual es la probabilidad de que sufran efectos secundarios más de 12 personas?*

La probabilidad de que sufran efectos secundarios 4 personas entre las 15 es:
$\binom{15}{4} 0.4^4 \times 0.6^{11} = \frac{15 \times 14 \times 13 \times 12}{4 \times 3 \times 2 \times 1} \times 0.025\,6 \times 3.\,628\,0 \times 10^{-3} = 0.126\,78$
la probabilidad de que sufran efectos secundarios más de 12 sería:
$\binom{15}{13} 0.4^{13} \times 0.6^2 + \binom{15}{14} 0.4^{14} \times 0.6 + \binom{15}{15} 0.4^{15} \times 0.6^0 = 2.\,789 \times 10^{-4}$

Ejercicio 33 *Una variable aleatoria X, que puede tomar valores dentro del espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$, tiene una función de probabilidad dada por la expresión $P(X = i) = \frac{i}{21}$, donde $i \in \Omega$.*

1. *Calcular el valor medio de esta variable*
2. *Calcular $P(1 \leq X \leq 3)$*
3. *Si extraemos una muestra de tres elementos (con reemplazamiento), ¿Cuál es la probabilidad de que, al menos uno de ellos, sea mayor que 3?*

1. $\overline{X} = 1 \times \frac{1}{21} + 2 \times \frac{2}{21} + 3 \times \frac{3}{21} + 4 \times \frac{4}{21} + 5 \times \frac{5}{21} + 6 \times \frac{6}{21} = 4.333$
2. $P(1 \leq X \leq 3) = \frac{1}{21} + \frac{2}{21} + \frac{3}{21} = 0.285\,71$
3. Pasamos al suceso contrario: Hallamos la probabilidad de que ninguno de los elementos sea mayor que tres

   Calculamos, en primer lugar, la probabilidad de que uno de los elementos extraidos no sea mayor que tres. Sólo se da esta circunstancia si extraemos 1 , 2 o 3.

   La probabilidad es por tanto: $P(X \leq 3) = \frac{1}{21} + \frac{2}{21} + \frac{3}{21} = 0.286$

   La probabilidad de que ninguno de los tres elementos extraidos sea sea mayor que tres será entonces:

   $0.286^3 = 0.023$.

   Por tanto la probabilidad del suceso contrario (al menos algun elemento es mayor que 3) es:

   $1 - [P(X \leq 3)]^3 = 1 - 0.023 = 0.977$

Ejercicio 34 *Se sabe que la probabilidad que tiene una persona de padecer una cierta enfermedad es 0.10. Para detectar si una persona padece esa enfermedad se le realiza una prueba médica. Esta prueba no es absolutamente fiable, ya qué si una persona está enferma no detecta la enfermedad en el 5% de los casos y si está sana la considera como enferma el 7% de los veces. Calcular*

1. *La probabilidad de que la prueba detecte la enfermedad en una persona.*
2. *La probabilidad de que una persona esté sana si la prueba médica le ha detectado la enfermedad.*
3. *La probabilidad de que una persona esté enferma si la prueba no se la ha detectado*

1. Sean S el suceso correspondiente a que una persona esté sana y E el suceso que se verifica cuando está enferma.

   Llamemos F al suceso correspondiente a que la prueba sea favorable (declare a la persona sana) y D el suceso correspondiente a que la prueba sea desfavorable (declare a la persona como enferma).

   En este caso los datos del problema son:

   $P(E) = 0.10$

   $p(F/E) = 0.05$

   $p(D/S) = 0.07$

   De ello se deduce:

   $P(E) = 0.10 \Longrightarrow P(S) = 0.90$

   $p(F/E) = 0.05 \Longrightarrow P(D/E) = 0.95$

   $p(D/S) = 0.07 \Longrightarrow P(F/S) = 0.93$

   Tendremos que calcular la probabilidad del suceso D. Se aplicará el teorema de la probabilidad total. Los sucesos S y E cumplen las condiciones:

   $S \cup E = \Omega$, $S \cap E = \Phi$, así que pueden tomar el papel de los sucesos A_i que aparecen en el enunciado de dicho teorema:

   $P(D) = P(S)P(D/S) + P(E)P(D/E) = 0.9 \times 0.07 + 0.10 \times 0.95 = 0.158$

2. Usaremos el teorema de Bayes:
   $P(S/D) = \frac{P(S)P(D/S)}{P(D)} = \frac{0.9 \times 0.07}{0.158} = 0.398\,73$

3. Tambien en este caso usaremos el teorema de Bayes
   $P(E/F) = \frac{P(E)P(F/E)}{P(F)} = \frac{P(E)P(F/E)}{P(S)P(F/S)+P(E)P(F/E)} = \frac{0.10 \times 0.05}{0.90 \times 0.93 + 0.10 \times 0.05} = 0.003$

Ejercicio 35 *Una compañia de seguros de automóviles clasifica sus polizas según el riesgo de accidente de cada uno de los vehículos asegurados. El 20% de las polizas son de alto riesgo (R_A), el 30% de riesgo medio (R_M) y el 50% de riesgo bajo (R_B). Se sabe que un vehículo con una póliza de alto riesgo tiene una probabilidad 0.3 de tener accidente (A) en el próximo año, una de riesgo medio una probabilidad de 0.1, y una de bajo riesgo una probabilidad de 0.001. Calcular*

1. *Calcular la probabilidad de que un vehículo asegurado, seleccionando al azar, sea de alto riesgo y tenga un accidente.*

2. *Calcular la probabilidad de que un vehículo asegurado, seleccionando al azar, tenga un accidente.*

3. *Si un vehículo determinado ha tenido un accidente, ¿Cual es la probabilidad de que tenga una poliza de alto riesgo?*

1. $P((R_A \cap A) = P(R_A).P(A/R_A) = 0.20 \times 0.3 = 0.06$

2. $P(A) = P(R_A).P(A/R_A) + P(R_M).P(A/R_M) + P(R_B).P(A/R_B) =$ $0.20 \times 0.3 + 0.30 \times 0.1 + 0.5 \times 0.001 = 0.090\,5$

3. $P(R_A/A) = \frac{P(R_A).P(A/R_A)}{P(A)} = \frac{0.20 \times 0.3}{0.090\,5} = 0.662\,98$

Ejercicio 36 *Durante el mes de Enero (20 días laborales), la probabilidad de que una persona pida un día de baja para asistir a una boda es 0.05. Si en una empresa hay 10 empleados.¿Cual es la probabilidad de que alguno de ellos pida baja por dicho motivo durante ese mes?*

Calculamos la probabilidad de que ninguno de ellos pida baja durante los 20 días. Hacemos el siguiente razonamiento:

La probabilidad de que una persona concreta no pida baja el primer día es 0.95.

La probabilidad de que ninguna persona pida baja ese día es 0.95^{10}.

La probabilidad de que esto último ocurra durante los 20 días laborables de Enero será $\left(0.95^{10}\right)^{20}$.

Por tanto, pasando al suceso contrario, la probabilidad de que haya alguna baja durante el mes de Enero es $1 - \left(0.95^{10}\right)^{20} = 0.999\,96$. Por tanto es prácticamente seguro que habrá alguna baja.

T. 3

Distribuciones Estadísticas

Ejercicio 37 *Dada la función:*

$$f(x) = \begin{cases} cx^2 & si\ \ 0 < x < 3 \\ 0 & en\, el\, resto \end{cases}$$

1. *Calcular c para que sea una función de probabilidad .*
2. *P(1<x<2)*
3. *Calcular la función de distribución correspondiente*

1. $\int\limits_0^3 cx^2\, dx = 9c = 1$. Por tanto $c = 1/9$
2. $\int\limits_1^2 \frac{1}{9}x^2\, dx = 0.259\,26$
3. La función de distribución es:

$$f(x) = \begin{cases} F(x) = 0 \quad para\ x \leq 0 \\ F(x) = P(X \leq x) = x^3/27 \qquad si\ \ 0 < x < 3 \\ F(x) = 1 \quad para\ x > 3 \end{cases}$$

Ejercicio 38 *Sea f(x)=* $\frac{1}{2}e^{-\frac{x}{2}}$ *para* $x >$ *0, la función de densidad de la variable que controla la duración un tipo de transistores (en cientos de horas).*

1. *Comprobar que es una función densidad*
2. *Hallar la función de distribución*
3. *Hallar la probabilidad de que dure entre 100 y 300 horas, usando la función de densidad y la de distribución*

4. *Halla la probabilidad de que uno de estos transistores dure más de 800 horas.*

1. Se verifican las propiedades que caracterizan a la función densidad de probabilidad: $\frac{1}{2}e^{-\frac{x}{2}} > 0$ y $\int\limits_0^\infty \frac{1}{2}e^{-\frac{x}{2}}dx = 1$.

2. $F(x) = \frac{1}{2}\int\limits_0^x e^{-\frac{x}{2}}dx = 1 - e^{-\frac{x}{2}}$

3. $P(1 < X < 3) = \frac{1}{2}\int\limits_1^3 e^{-\frac{x}{2}}dx = e^{-\frac{3}{2}} + e^{-\frac{1}{2}} = -0.2231 + 0.6065 = 0.383$

   $P(1 < X < 3) = P(X < 3) - P(X < 1) = (1 - e^{-\frac{3}{2}}) - (1 - e^{-\frac{1}{2}}) = 0.383$

4. $P(X > 8) = 1 - P(X \leq 8) = 1 - (1 - e^{-4}) = e^{-4} = 0.0182$

Ejercicio 39 *El encargado de una gasolinera, recogiendo los datos de ventas durante bastantes semanas, ha llegado a la conclusión de que la demanda semanal de gasolina (en Kl) sigue proximadamente la distribución dada por la función de densidad*

$$f(x) = \begin{cases} x & si\ \ 0 \leq x \leq 1 \\ \frac{1}{2} & si\ \ 1 < x \leq 2 \\ 0 & en\, el\, resto \end{cases}$$

1) Calcular la probabilidad de que en una semana se demanden entre 0.5 y 1 Kl.
2) Entre 0.9 y 1.1 Kl.
3) Más de 1500 l.
4) La demanda semanal esperada

1. $\int\limits_{0.5}^1 x\,dx = \left.\frac{x^2}{2}\right|_{0.5}^1 = 1/2 - \frac{0.5^2}{2} = 0.375$

2. $\int\limits_{0.9}^1 x\,dx + \int\limits_1^{1.1} \frac{1}{2}\,dx = \left.\frac{x^2}{2}\right|_{0.9}^1 + \left.\frac{1}{2}x\right|_1^{1.1} = 1/2 - \frac{0.9^2}{2} + \frac{1.1}{2} - \frac{1}{2} = 0.145$

3. $\int\limits_{1.5}^2 \frac{1}{2}dx = 1/4$

4. $\int\limits_0^1 x.x\,dx + \int\limits_1^2 x\frac{1}{2}\,dx = \left.\frac{x^3}{3}\right|_0^1 + \left.\frac{x^2}{4}\right|_1^2 = 1/3 + 3/4 = 1.08$ Kl. es la media de ventas semanales.

Ejercicio 40 *Sea X una variable aleatoria N(3, 2), calcular:* $P(x \leq 5)$, $P(x > 3)$, $P(0.4 < x < 3.2)$

$P(x \leq 5) = P\left(\frac{x-3}{2} \leq \frac{5-3}{2}\right) = P(z \leq 1) = 0.8413$
$P(x > 3) = P\left(\frac{x-3}{2} > \frac{3-3}{2}\right) = P(z > 0) = 1 - P(z \leq 0) = 0.5$
$P(0.4 < x < 3.2) = P\left(\frac{0.4-3}{2} < \frac{x-3}{2} < \frac{3,2-3}{2}\right) = P(-1.3 < x < 0.1) =$
$P(z < 0.1) - P(z - 1,3) = P(z < 0.1) - (1 - P(z < 1.3))$
$= 0.5398 + 0.9032 - 1 = 0.4430$

Ejercicio 41 *La demanda por segundo de agua de una estación de bombeo tiene una media de 100 m*^{3}*/seg. y sigue una distribución exponencial*

1) Calcular la probabilidad de que el agua demandada en un cierto segundo sea superior a 200m3

2) Se quiere que la demanda sea atendida al menos en el 99% de los casos. ¿Cuánta agua ha de estar disponible?

1. La función de densidad en cientos de m^3 es $f(x) = e^{-x}$.

2. $P(x > 2) = \int\limits_2^\infty e^{-x}dx = e^{-2} = 0.135$

3. $\int\limits_0^t e^{-x}dx = 0.99; \quad \Longrightarrow 1 - e^{-t} = 0.99, \quad e^{-t} = 0.01; \quad t = 4.60.$

   Luego la cantidad de agua disponible ha de ser de 460 m3/seg

Ejercicio 42 *Un lote de piezas contiene un 20% de defectuosas. Un cliente decide comprar el lote si tomando 100 piezas de éste elegidas al azar, como máximo 12 son defectuosas. Calcular la probabilidad de aceptar el lote.*

Usamos una $B(100, 0.2)$
$P(x \leq 12) =$
$= \begin{pmatrix} 100 \\ 0 \end{pmatrix} 0.2^0.0.8^{100} + \begin{pmatrix} 100 \\ 1 \end{pmatrix} 0.2^1.0.8^{99} + ... + \begin{pmatrix} 100 \\ 12 \end{pmatrix} 0.2^{12}.0.8^{88}$

Como el cálculo es complicado vamos a aproximarlo con una normal. Se cumplen las condiciones para que sea válida la aproximación:

$np = 20 > 5,\ nq = 80 > 5,\ p = 0.20 > 0.05,\ q = 0.80 > 0.05$

Aproximamos la binomial con una distribución $N(np, \sqrt{npq}) = N(20, 4)$

Usando la corrección por continuidad resulta:

$P(x < 12.5) = P\left(\frac{x-20}{4} < \frac{12.5-20}{4}\right) = P(z < -1.875) = 0.0304$. Hay pocas probabilidades de aceptar un lote. Sólo se aceptaría el 3.04% de estos lotes.

Ejercicio 43 *Se supone que el número de automóviles que pasan por un cruce de carretera en 5 mínutos sigue una distribución de Poisson de media 20 . Calcular:*

1. *Probabilidad de que pasen menos de 2 automóviles durante 5 minutos de observación.*

2. *Probabilidad de que pasen menos de 20.*

1. $P(0) + P(1) = \frac{20^0}{0!}e^{-20} + \frac{20^1}{1!}e^{-20} = 4.328\,4 \times 10^{-8}$.

2. $P(0)+P(1)+.....+P(18)+P(19) = \frac{20^0}{0!}e^{-20}+\frac{20^1}{1!}e^{-20}+....+\frac{20^{18}}{18!}e^{-20}+\frac{20^{19}}{19!}e^{-20}$

Usaremos la aproximación de la Poisson por la normal de media 20 y de varianza 20:
$P(x < 19.5) = P\left(\frac{x-20}{\sqrt{20}} < \frac{19.5-20}{\sqrt{20}}\right) = P(z < -0,118) = 1 - P(z < 0.118) = 1 - 0.5438 = 0.4562$.

Ejercicio 44 *Se sabe que el número de matrimonios que se registran cada mes en una ciudad española sigue una distribución normal de media 124 y desviación típica 8. Calcular:*

1. *La probabilidad de que un cierto mes el número de matrimonios registrados esté comprendido entre 112 y 130.*

2. *La probabilidad de que la media mensual de matrimonios, obtenida computando los matrimonios registrados durante cada uno de los 12 meses del año 2005, esté entre 122 y 126.*

1. $P(112 < X < 130) = P(X < 130) - P(X < 112) = P(z < \frac{130-124}{8}) - P(z < \frac{112-124}{8}) = P(z < 0.75) - P(z < -1.5)$

   $= 0.773\,37 - 6.680\,7 \times 10^{-2} = 0.706\,56$

2. la media se distribuye como una normal de media 124 y de desviación típica $\frac{8}{\sqrt{12}} = 2.31$

   $P(122 < \overline{X} < 126) = P(z < \frac{126-124}{2.31}) - P(z < \frac{122-124}{2.31}) = P(z < 0.865\,8) - P(z < -0.865\,8)$

   $= P(z < 0.8658) - P(z < -0.8658) = 0.806\,7 - 0.193\,3 = 0.613\,4$

Ejercicio 45 *La variable aleatoria X sigue una distribución de probabilidad cuya función de densidad es la siguiente*

$$f(x) = \begin{cases} \frac{x}{8} & si\ 0 \leq x \leq 4 \\ 0 & en\ el\ resto \end{cases}$$

Se pide

1. *Calcular el valor medio de X*
2. *La mediana de X*
3. *La función de distribución de X*
4. *$P(2 \leq X \leq 3)$*
5. *Si extraemos una muestra de tres elementos, ¿Cuál es la probabilidad de que, al menos uno de ellos, sea mayor que 2?*

1. $\mu = \int_0^4 x\frac{x}{8}dx = \frac{8}{3} = 2.6667$

2. $\int_0^m \frac{x}{8}dx = 0.5 \Longrightarrow \int_0^m \frac{x}{8}dx = \frac{1}{16}m^2 = 0.5$. Por lo tanto $m^2 = 8$ y $m = \sqrt{8} = 2.8284$

3. $$F(x) = \begin{cases} 0 & si\ x < 0 \\ \int_0^x \frac{x}{8}dx = \frac{x^2}{16} & si\ 0 \leq x \leq 4 \\ 1 & si\ x > 4 \end{cases}$$

4. $P(2 \leq X \leq 3) = \int_2^3 \frac{x}{8}dx = 0.3125$

5. Calcularemos la probabilidad del suceso contrario: Que ninguno de los tres elementos sea mayor que 2.

   Para ello calculamos, en primer lugar la probabilidad de que uno de ellos no sea mayor que 2: $1 - \frac{2^2}{16} = 0.75$. Entonces la probabilidad de que ninguno de los tres sea mayor que 2 es $0.75^3 = 0.42188$. Por tanto la probabilidad de que al menos uno de los elementos sea mayor que 2 será $1 - 0.42188 = 0.57812$.

Ejercicio 46 *En medicina es importante la dosis recomendada para un medicamento. Un laboratorio comercializa unos comprimidos cuyo peso sigue una distribución normal con un peso medio de 3 gramos con una desviación típica de 0.05 gramos.*

1. *Calcular la probabilidad de que un comprimido pese más de 3.025 gramos.*
2. *Un comprimido se considera defectuoso si su peso se aparta de la media en más de 0.08 gramos. Calcular la probabilidad de que un comprimido sea defectuoso.*
3. *Estos comprimidos se venden en cajas de 10 unidades. Si una caja contiene más de dos unidades defectuosas se retiran del mercado. ¿Qué porcentaje de cajas se retirarán del mercado?*

1. $P(x > 3.025) = P(\frac{x-3}{0.05} > \frac{3.025-3}{0.05}) = P(z > 0.5) = 1 - p(z < 0.5) = 0.308\,54$
2. $P(2.92 < x < 3.08) = P\left(\frac{2.92-3}{0.05} < z < \frac{3.08-3}{0.05}\right) = P\left(-1.6 < z < 1.6\right) = P(z < 1.6) - P(z < -1.6) = 2P(z < 1.6) - 1 = 2 \times 0.945\,2 - 1 = 0.890\,4.$

   La probabilidad de que sea defectuoso es $1 - 0.890\,4 = 0.109\,6$
3. $1 - \binom{10}{0} 0.1096^0 \times 0.890\,4^{10} - \binom{10}{1} 0.1096^1 \times 0.890\,4^9 - \binom{10}{2} 0.1096^2 \times 0.890\,4^8 = 8.767\,4 \times 10^{-2}$

   El 8.77% de las cajas habrá que retirarlas del mercado.

Ejercicio 47 *El intervalo de tiempo promedio entre la llegada de dos clientes consecutivos a la caja de un supermercado es de 12 seg en un día de promoción. Con motivo de esta promoción, se cuentan los clientes que se van incorporando a la cola y se hace entrega de un pequeño obsequio a los que les correspondería un número multiplo de 50.*

1. *¿Cual es la distribución que rige el intervalo de tiempo entre la entrega de dos obsequios consecutivos?*
2. *¿Cuál es su media y su desviación típica?*

1. El tiempo que pasa hasta que llegan 50 clientes consecutivos se rige por una distribución de Erlang:

$$f(x; n, \lambda) = \frac{\lambda}{(n-1)!} (\lambda x)^{n-1} e^{-\lambda x} \qquad \textit{para } x > 0$$

   En este caso $n = 50$, $\lambda = \frac{1}{12}$ llegadas por segundo. Por tanto la función de densidad de probabilidad que rige el intervalo de tiempo transcurrido entre la entrega de dos premios consecutivos será

$$\frac{\frac{1}{12}}{49!} \left(\frac{x}{12}\right)^{49} e^{-\frac{x}{12}} \qquad \textit{para } x > 0$$

2. La *media* de esta distribución es $\frac{n}{\lambda}$ y la *varianza* es $\frac{n}{\lambda^2}$. Por tanto el valor medio será $50 \times 12 = 600$ seg., es decir 10 minutos y la desviación típica $\sqrt{50 \times 12^2} = 84.85$ seg

Ejercicio 48 *Una caja contiene 15 tornillos de los cuales 5 son defectuosos Calcular la distribución de probabilidad que corresponde a la variable aleatoria correspondiente al número de tornillos defectuosos obtenidos al sacar 4 tornillos de la citada caja*

Sea i el número de tornillos defectuosos, que es la variable aleatoria asociada al experimento:

$$P(i=0) = \frac{\binom{5}{0}\binom{10}{5}}{\binom{15}{5}} = 8.3916 \times 10^{-2}$$
$$P(i=1) = \frac{\binom{5}{1}\binom{10}{4}}{\binom{15}{5}} = 0.34965$$
$$P(i=2) = \frac{\binom{5}{2}\binom{10}{3}}{\binom{15}{5}} = 0.3996$$
$$P(i=3) = \frac{\binom{5}{3}\binom{10}{2}}{\binom{15}{5}} = 0.14985$$
$$P(i=4) = \frac{\binom{5}{4}\binom{10}{1}}{\binom{15}{5}} = 0.01665$$

Observamos que la suma de las probabilidas es 1 (salvo errores númericos de truncanmiento).

$8.3916 \times 10^{-2} + 0.34965 + 0.3996 + 0.14985 + 1.665 \times 10^{-2} = 0.99967$

La distribución es hipergeométrica. La expresión general es

$$P(i=x) = \frac{\binom{5}{x}\binom{10}{5-x}}{\binom{15}{5}}, \qquad x = 0,1,2,3,4$$

Ejercicio 49 *La probabilidad de que cuando llames por telefono a cierta oficina de información de RENFE esté comunicando es 0.40.*

1. *Calcular la probabilidad de poder comunicar al primer intento*
2. *Calcular la probabilidad de no poder hacerlo hasta el segundo intento*
3. *Calcular la probabilidad de no poder hacerlo hasta el tercer intento*
4. *Calcular la probabilidad de poder comunicar antes del quinto intento*
5. *Calcular la probabilidad de no poder comunicar hasta despues del quinto intento.*

1. La probabilidad de poder comunicarse al primer intento es: $1 - 0.40 = 0.60 = 0.6$

2. La probabilidad de no poder hacerlo hasta el segundo intento es: 0.40×0.60

3. La probabilidad de no poder hacerlo hasta el tercer intento es $(0.40)^2\times 0.60$

4. La probabilidad de poder comunicarse antes del quinto intento

   $0.6+(0.40)\times 0.60+(0.40)^2\times 0.60+(0.40)^3\times 0.60=0.974\,4$

5. La probabilidad de no poder comunicarse hasta despues del quinto intento es:

   $1-\Big(0.6+(0.40)\times 0.60+(0.40)^2\times 0.60+(0.40)^3\times 0.60+(0.40)^4\times 0.60\Big)$

   El modelo es la distribución geométrica.

Ejercicio 50 *El tiempo en horas hasta que se produce un fallo de un tipo de componentes electrónicos sigue una distribución de Weibull de parámetros* γ *=50,* $\beta=\frac{1}{3}$ *,*$\eta=$ *100. Calcular*

1. *La probabilidad de que uno de estos elementos falle antes de 300 horas.*

2. *El tiempo medio hasta el fallo de este tipo de dispositivos.*

1. $F(t)=1-e^{--\left(\frac{t-\gamma}{\eta}\right)^{\beta}}=1-e^{-\left(\frac{300-50}{100}\right)^{\frac{1}{3}}}=1-e^{-\left(\frac{300-50}{100}\right)^{\frac{1}{3}}}=1-e^{-(2.\,5)^{\frac{1}{3}}}=1-e^{-1.\,357\,2}=1-\frac{1}{e^{1.3572}}=0.742\,62$

2. $\mu=\gamma+\eta\,\Gamma\left(1+\frac{1}{\beta}\right)=50+100\times\ \Gamma\left(4\right)=650.0$

Ejercicio 51 *Se tira un dado hasta obtener tres veces el cinco.*

1. *Calcular la probabilidad de que esto ocurra a la septima tirada.*

2. *Calcular la probabilidad de que esto ocurra antes de la septima tirada.*

1. $\binom{6}{2}\frac{1}{6}^2\frac{5}{6}^4\frac{1}{6}=3.\,349\,0\times 10^{-2}$

2. $\binom{2}{2}\frac{1}{6}^2\frac{5}{6}^0\frac{1}{6}+\binom{3}{2}\frac{1}{6}^2\frac{5}{6}^1\frac{1}{6}+\binom{4}{2}\frac{1}{6}^2\frac{5}{6}^2\frac{1}{6}+\binom{5}{2}\frac{1}{6}^2\frac{5}{6}^3\frac{1}{6}=6.\,228\,6\times 10^{-2}$

Ejercicio 52 *Se ha estimado que el número de enfermos atendidos en un consultorio médico cada 10 minutos se distribuye según una ley de Poisson con una media de 3.8. Calcular la probabilidad de que en un intervalo de 10 minutos sean atendidos:*

1. *Ningún enfermo,*

2. Un enfermo

3. Al menos dos enfermos

1. $P\{X=0\} = \frac{3.8^0}{0!}e^{-3.8} = 2.2371 \times 10^{-2}$

2. $P\{X=1\} = \frac{3.8^1}{1!}e^{-3.8} = 8.5009 \times 10^{-2}$

3. $1 - P(X=0) - P(X=1) = 1 - 2.2371 \times 10^{-2} - 8.5009 \times 10^{-2} = 0.89262$

Ejercicio 53 *Supongamos que el ingreso mensual de un camarero es una variable aleatoria cuya función de densidad está determinada por:*

$$f(x) = ke^{-\frac{x}{800}}, \quad x > 0$$

1. ¿Cuánto debe valer k para que sea una función de densidad?

2. Obtener la función de distribución.

3. Calcular la probabilidad de que el ingreso mensual exceda el ingreso promedio.

4. Determinar los ingresos medianos y el recorrido interdecil.

1. Para que f(x) sea una función de densidad se ha de cumplir:

$$\int_0^\infty ke^{-\frac{x}{800}}dx = 1$$

así que

$$k = \frac{1}{\int_0^\infty e^{-\frac{x}{800}}dx} = \frac{1}{-800e^{-\frac{x}{800}}\Big|_0^\infty} = -\frac{1}{0-800} = \frac{1}{800}$$

2. La función de distribución es:

$$F(x) = \int_0^x \frac{1}{800}e^{-\frac{x}{800}} = 1 - e^{-\frac{x}{800}}, \quad x > 0$$

Calculamos los ingresos medios:

$$\mu = \int_0^\infty x\frac{1}{800}e^{-\frac{x}{800}}dx$$

realizando en primer lugar por partes la correspondiente integral definida:

$$\int x\frac{1}{800}e^{-\frac{x}{800}}dx = -800e^{-\frac{1}{800}x} - xe^{-\frac{1}{800}x}$$

y aplicando la regla de Barrow obtenemos:

$$\mu = \int_0^\infty x\frac{1}{800}e^{-\frac{x}{800}}dx = 0-(-800) = 800$$

3. $P(X > 800) = 1 - P(X \leq 800) = 1 - F(800) = e^{-\frac{800}{800}} = 0.367\,88$

4. Para calcular la mediana tenemos en cuenta que si M es el ingreso mediano se ha de verificar que $P(X \leq M) = 0.5$:

$$\int_0^M \frac{1}{800}e^{-\frac{x}{800}}dx = \left. -800e^{-\frac{x}{800}}\right|_0^M = -e^{-\frac{M}{800}} + 1 = 0.5$$

De esta expresión se deduce que

$$e^{-\frac{M}{800}} = 0.5; \quad \ln 0.5 = .-\frac{M}{800}$$

Despejando resulta $M = 554.\,52$

El recorrido interdecílico es la diferencia entre el noveno y primer decil: $D_9 - D_1$. Calculamos, en primer lugar el noveno decil:

$F(D_9) = 0.9$, por tanto $-e^{-\frac{D_9}{800}}+1 = 0.9$; $\;e^{-\frac{D_9}{800}} = 0.1$; $\;\ln 0.1 = -\frac{D_9}{800}$; así que $D_9 = 1842.1$

Hallamos ahora el primer decil:

$F(D_1) = 0.10$; $\;-e^{-\frac{D1}{800}} + 1 = 0.1$; $\;e^{-\frac{D_9}{800}} = 0.9$; $\;, \ln 0.9 = -\frac{D_1}{800}$, por tanto $D_1 = 84.288$.

$D_9 - D_1 = 1842.\,1 - 84.\,288 = 1757.8$ es el recorrido interdecílico.

T. 4

Simulación y teorema central del límite

Ejercicio 54 *Generar 100 valores con el método de generación de números aleatorios de algún paquete estadístico y comprobar la calidad del algoritmo usando esta muestra:*

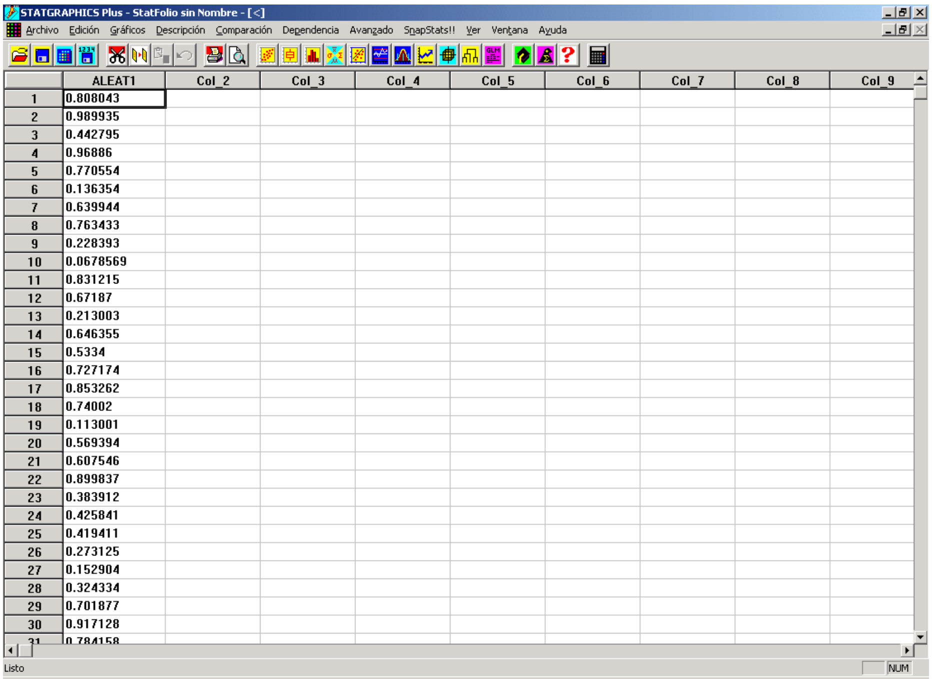
1. *Por medio del test Chi-cuadrado para comprobar el ajuste de la muestra a la distribución U[0,1].*

2. *Usando las funciones de autocorrelación muestral para comprobar la independencia de los valores muestrales.*

Se generará una muestra con 100 elementos con cualquier paquete estadístico. En este caso, describimos la forma de realizar esta operación con Statgraphics Plus 5.1. Para generar la muestra de 100 elementos requerida se entra en el procedimiento **Descripción** del Menú Principal:

$$\textit{Descripción} \rightarrow \textit{Distribucion}$$

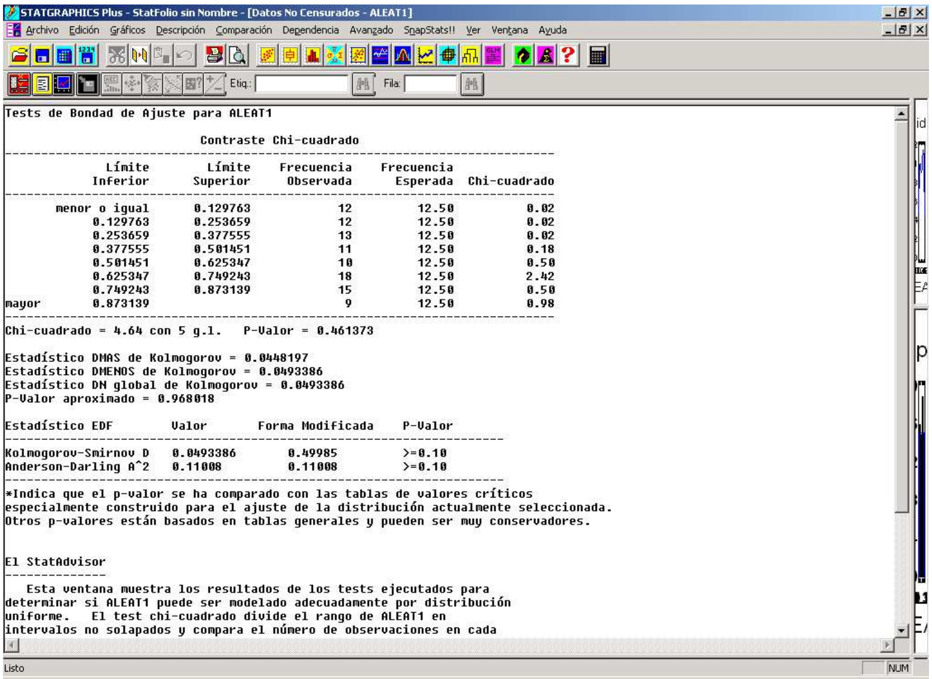
Seleccionamos la opción **Uniforme**. Posteriormente, situandonos sobre cualquiera de las ventanas emergentes y usando el botón derecho del ratón, seleccionamos **Opciones del Análisis** para fijar los extremos, 0 y 1, de la distribución. Despues, seleccionamos dentro de la ventana del procedimiento el icono amarillo, llamado **Opciones tabulares,** y en la ventana que se despliega **Numeros aleatorios**. De esta forma se generan 100, valor por defecto, numeros aleatorios con la distribución seleccionada. En este caso uniforme [0,1]. Para verlos en el fichero de datos hay que grabarlos. Esta operación se realiza pulsando un icono que representa un disquete, y seleccionando posteriormente la casilla correspondiente. Se puede dejar el nombre que nos ofrece el programa, ALEAT1, para la variable que contenga los valores generados. De esta forma, abriendo el fichero de datos podremos ver estos valores. En la siguiente figura se muestran los primeros valores generados. Es de notar que si se repite la experiencia no se repiten los mismos valores, ya que se pretende imitar un compontamiento debido al azar.



1. Para realizar el test Chi cuadrado de bondad de ajuste a una uniforme U[0,1] selecionamos de nuevo

   $$Descripción \rightarrow Distribucion$$

   pero ahora seleccionamos **Ajuste de distribuciones (datos no censurados)**, introduciendo la variable ALEAT1 en la casilla de datos. La distribución que ajusta el programa por defecto es la normal así que hay que cambiarla. Para ello, y como anteriormente, seleccionamos **Opciones del Análisis y** elegimos la distribución uniforme. No se pueden fijar ahora los extremos, 0 y 1, de la distribución. Esto es una límitación del programa. Observar si el programa ha elegido al menos valores próximos a estos extremos. Mirando la salida del programa comprobamos que aparecen los resultados de diferentes test. Fijandonos en la prueba Chi- cuadrado miramos el P-value. En está prueba el resultado, como puede apreciarse en la siguiente figura es 0.461373. Si este valor es mayor que 0.05, tal como ocurre en este caso, aceptamos que los datos se ajustan razonablemente a una distribución uniforme [0,1].



STATGRAPHICS Plus - StatFolio sin Nombre - [Datos No Censurados - ALEAT1]

Archivo Edición Gráficos Descripción Comparación Dependencia Avanzado SnapStats!! Ver Ventana Ayuda

Etiq.: Fila:

Tests de Bondad de Ajuste para ALEAT1

Contraste Chi-cuadrado

	Límite Inferior	Límite Superior	Frecuencia Observada	Frecuencia Esperada	Chi-cuadrado
	menor o igual	0.129763	12	12.50	0.02
	0.129763	0.253659	12	12.50	0.02
	0.253659	0.377555	13	12.50	0.02
	0.377555	0.501451	11	12.50	0.18
	0.501451	0.625347	10	12.50	0.50
	0.625347	0.749243	18	12.50	2.42
	0.749243	0.873139	15	12.50	0.50
mayor	0.873139		9	12.50	0.98

Chi-cuadrado = 4.64 con 5 g.l.   P-Valor = 0.461373

Estadístico DMAS de Kolmogorov = 0.0448197
Estadístico DMENOS de Kolmogorov = 0.0493386
Estadístico DN global de Kolmogorov = 0.0493386
P-Valor aproximado = 0.968018

Estadístico EDF	Valor	Forma Modificada	P-Valor
Kolmogorov-Smirnov D	0.0493386	0.49985	>=0.10
Anderson-Darling A^2	0.11008	0.11008	>=0.10

*Indica que el p-valor se ha comparado con las tablas de valores críticos especialmente construido para el ajuste de la distribución actualmente seleccionada. Otros p-valores están basados en tablas generales y pueden ser muy conservadores.

El StatAdvisor

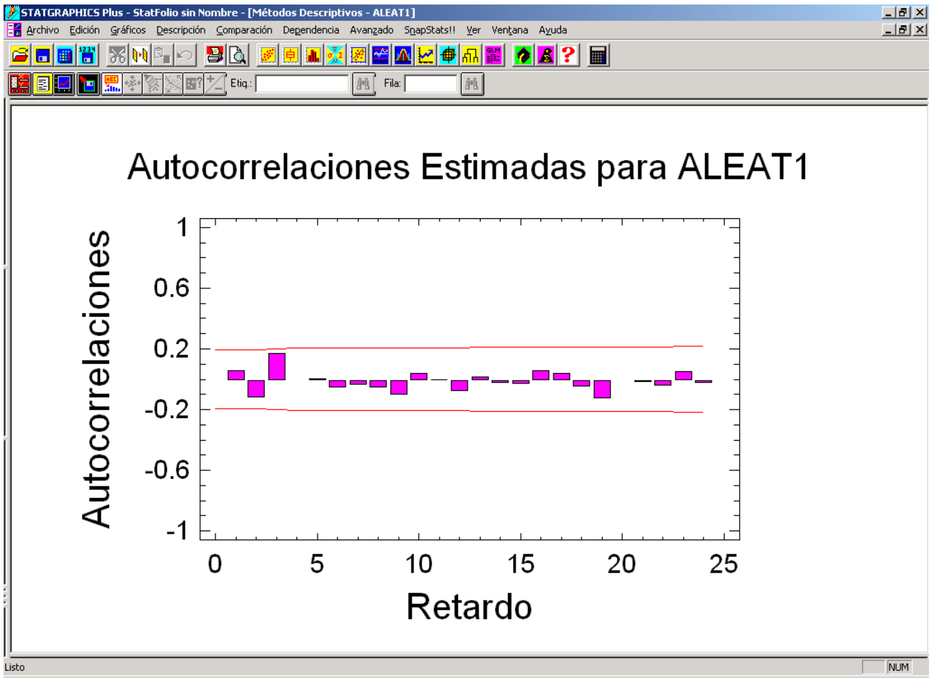
Esta ventana muestra los resultados de los tests ejecutados para determinar si ALEAT1 puede ser modelado adecuadamente por distribución uniforme. El test chi-cuadrado divide el rango de ALEAT1 en intervalos no solapados y compara el número de observaciones en cada

Listo NUM

2. Esta segunda prueba trata de revisar si los datos muestrales no guardan relaciones de dependencia entre sí. Con Statgraphics se accede a esta prueba siguiendo la siguiente secuencia desde el menu principal:

$$\textit{Avanzado} \rightarrow \textit{Analisis de series temporales} \rightarrow \textit{descriptivo}$$

Seleccionando la variable, y aceptando el análisis, nos aparecen por defecto varias ventanas. Aunque el programa permite acceder a otras pruebas de independencias, un procedimiento sencillo se basa en observar la gráfica segunda, la de las autocorrelaciones. Si ninguno de los rectángulos, como ocurre en este caso según consta en la siguiente figura, sobrepasa las lineas paralelas aceptamos que los valores muestrales son independientes entre sí.



Ejercicio 55 *1. Construir un generador de números aleatorios usando el método congruencial, basado en la relación*

$$X_{i+1} \equiv (aX_i + c)\ (mod\ \ m)$$

$$c = 7,\ a = 5,\ m = 2^{16}.$$

2. ¿Cuántos números distintos genera este algoritmo?

3. Usando este algoritmo anterior genera valores procedentes de una distribución U[3,6].

1. El algoritmo puede constar de los siguientes pasos:

   (a) Se decide el valor de partida del algoritmo, que llamaremos $SEMILLA$ (número natural) y el número de valores que se desea generar que llamamos NV (número natural). Definimos un contador.

   (b) $contador = 0,\ c = 7,\ a = 5,\ m = 65536$

   (c) $X_0 = SEMILLA,\ N = NV$

   (d) $contador = contador + 1,\ X = a \times X_0 + c$

(e) X = resto de la división de X entre m

(f) $rand = X/m$. Escribir $rand$

(g) $X_0 = X$. Si $contador = NV$, STOP

(h) Ir al paso d)

Realizamos un ejemplo de aplicación del algoritmo. Generaremos tres numeros aleatorios en el intervalo [0,1]

$Semilla = 3512$ $NV = 3$	contador =1	Contador=2	Contador=3
X_0	3512	17567	22306
$X = 5X_0 + 7$	17567	87842	111537
X= resto de la división x y 65536	17567	22306	46001
rand	0.26805	0.34036	0.70192

2. Este algoritmo es de periodo completo, genera $2^{16} = 65536$.números distintos. Esto es así porque se cumplen las condiciones apropiadas para que el periodo sea completo:

   a) c y m son primos entre sí.

   b) $a \equiv 1(\text{mod } g)$ Siendo g cualquier factor primo de m. En este caso $g = 2$.

   c) $a \equiv 1$ (mod 4) si m es múltiplo de 4.

3. Se pueden emplear los valores generados en el intervalo $[0, 1]$. Para transformarlos a valores en el intervalo $[3, 6]$ realizamos un cambio de origen y de escala Si $rand \in [0, 1]$, entonces $rand2 = 3 + 3 \times rand$ pertenece al intervalo $[3, 6]$.

   Por ejemplo, usando los tres valores generados en el apartado anterior podemos generar 3 valores en el intervalo $[3, 6]$:

   rand

   $3 + 3 \times 0.26805 = 3.8042$

   $3 + 3 \times 0.34036 = 4.0211$

   $3 + 3 \times 0.70192 = 5.1058$

Ejercicio 56 *Simula con el ordenador la tirada de tres dados.*

Como en cada dado la probabilidad de cualquier cara es 1/6 puede hacerse con un generador de numeros aleatorios $U[0, 1]$ haciendo la siguiente correspondencia: Si el número generado es menor que $\frac{1}{6}$ le asignamos la puntuación 1 si rand$<\frac{1}{6}$.

Si el número generado es mayor o igual que $\frac{1}{6}$, pero menor que $\frac{2}{6}$ le asignamos la puntuación 2 y así sucesivamente.

Si usamos los números obtenidos en el ejercicio 1, 0.26805, 0.34036, 0.70192, obtendríamos para los tres dados las puntuaciones $2, 3, 4$. Si quisieramos realizar otra tirada de tres dado generaríamos otros tres números aleatorios.

Ejercicio 57 *Implementa un algoritmo que genere valores controlados por una distribución exponencial.*

Se puede utilizar el método de la transformación inversa:

La función de densidad es $f(x) = \lambda.exp(-\lambda x)$:

Si $\eta \in U[0,1]$ es el número generado por una U[0,1]

$$\eta = F(x) = \int_0^x \lambda.exp(-\lambda x)dx = 1 - exp(-\lambda x)$$

Despejando x se obtiene:

$$x = -\frac{1}{\lambda}L(1-\eta).$$

Como $1-\eta$ sigue la misma distribución que η , se puede calcular :

$$x = -\frac{1}{\lambda}L\eta$$

Si $\eta \in U[0,1]$, entonces x se distribuye como una exponencial de parámetro λ.

La siguiente subrutina FORTRAN implementa este método. Esta subrutina necesita otra que no está incluida que sirva para generar números aleatorios uniformemente distribuidos en [0,1]. Esta subrutina es utilizada por medio de la llamada: CALL RND(SEM,U). SEM es la Semilla y U el número aleatorio generado por RND.

```
!*********SUBRUTINA DISTRIB.EXPONENCIAL*** EXPO ******
SUBROUTINE EXPO(NPUN,SEM)
IMPLICIT NONE
INTEGER*4 NPUN,SEM,N
EXTERNAL RND
REAL*8 U,X,LANDA
OPEN(UNIT=4,FILE='EXPO.TXT',STATUS="REPLACE")
PRINT*,'INTRODUZCA EL PARAMETRO DE LA EXPONENCIAL'
READ*,LANDA
DO N=1,NPUN
    CALL RND(SEM,U)
```

```
      X=-1.D0/LANDA*DLOG(U)
      WRITE(4,*)X
      PRINT*,X
   END DO
   CLOSE(4)
   END SUBROUTINE
```

Ejercicio 58 *Implementa un algoritmo que genere valores controlados por una distribución binomial.*

La siguiente subrutina BINO asocia a cada valor de la variable aleatoria U∈U[0,1] un valor para la variable aleatoria binomial X, de modo que se verifique: $\sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i} \leq U \leq \sum_{i=0}^{x+1} \binom{n}{i} p^i q^{n-i} \quad x = 0, 1, \cdots, n$

```
   !*******SUBRUTINA DISTRIB.BINOMIAL**** BINO *********
   SUBROUTINE BINO(NPUN,SEM)
   IMPLICIT NONE
   INTEGER*4 NPUN,SEM,N,NN,I
   EXTERNAL RND
   REAL*8 U,X,P,Q,P2,P3,A,A1,A2,A3
   OPEN(UNIT=4,FILE='BINOMIAL.TXT',STATUS="REPLACE")
   PRINT*,'INTRODUZCA LOS VALORES DE N Y P'
   READ*,N,P
   Q=1-P
   DO WHILE (NN<>NPUN)
      X=0
      P2=Q**N
      CALL RND(SEM,U)
      DO WHILE (U>P2)
      A1=1
      A2=1
      A3=1
   !cálculo de factoriales
   IF(N.EQ.0)THEN
      A1=1.0D00
   ELSE
      DO I=1,N
   A1=A1*I
   ENDDO
      ENDIF
      IF(X+1.EQ.0)THEN
      A2=1.0D00
```

```
ELSE
     DO I=1,X+1
A2=A2*I
ENDDO
     ENDIF
     IF(N-X-1.EQ.0)THEN
     A3=1.0D00
ELSE
     DO I=1,N-X-1
A3=A3*I
ENDDO
     ENDIF
          P2=P2+A1/(A2*A3)*P**(X+1)*Q**(N-X-1)
          X=X+1
     ENDDO
     WRITE(4,*)X
     PRINT*,X
     NN=NN+1
END DO
CLOSE(4)
END SUBROUTINE
```

Ejercicio 59 *Teniendo en cuenta que la suma de variables exponenciales es una variable de Erlang, implementar un algoritmo para generar valores que se rijan por una distribución de Erlang.*

Las siguientes sentencias de la subrutina ERLANG realizan la suma de "NERLANG" variables exponenciales. A cada valor de I se añade un nuevo factor a u_2, de modo que. al final del DO, u_2 es el producto de "NERLANG" variable U [0,1], y por tanto X=-1.D0*LANDA*DLOG(U2) tomará el valor $-\lambda \ln(u_1 \times u_2 \times \cdots u_{nerlang}) = -\lambda \ln(u_1) - \lambda \ln(u_2) - \cdots \lambda \ln(u_{nerlang})$, que es la suma de "NERLANG" exponenciales.

```
DO I=1,NERLANG
               CALL RND(SEM,U1)
               U2=U2*U1
          ENDDO
          X=-1.D0*LANDA*DLOG(U2)
          WRITE(4,*)X
```

```
!*******SUBRUTINA DISTRIB ERLANG************************
SUBROUTINE ERLANG(NPUN,SEM)
```

```
    INTEGER*4 NPUN,SEM,N
    EXTERNAL RND
    REAL*8 X,NERLANG,LANDA,U1,U2
    OPEN(UNIT=4,FILE='ERLANG.TXT',STATUS="REPLACE")
    PRINT*,'INTRODUZCA LOS PARAMETROS NERLANG(POSITIVO
Y ENTERO) Y LANDA '
    READ*,NERLANG,LANDA
            U2=1
            N=0
            DO WHILE (N<>NPUN)
                DO I=1,NERLANG
                    CALL RND(SEM,U1)
                    U2=U2*U1
                ENDDO
                X=-1.D0*LANDA*DLOG(U2)
                WRITE(4,*)X
                PRINT*,X
                N=N+1
                U2=1
            END DO
    CLOSE(4)
    END SUBROUTINE
```

Ejercicio 60 *Una variable que se distribuye uniformemente entre 7500 y 10500. hallar la probabilidad de que la media de una muestra de 1000 elementos de esta variable sea mayor que 9050.*

1. La media de las rentas tiene, usando el T.C.L. en la forma de Lindeberg-Levy, una distribución que es aproximadamente normal, cuya media es la de la distribución de partida, la uniforme U[7500, 10500], y desviación típica es la desviación típica de dicha distribución uniforme dividida por la raiz cuadrada de los elementos de la muestra considerada.

   La media y la desviación típica de la uniforme son:

   $\mu = \frac{7500+10500}{2} = 9000$, $\sigma = \frac{(10500-7500)}{\sqrt{12}} = 866.03$

   Por tanto la variable aleatoria media de las muestras de 1000 elementos de la distribución uniforme ,$\overline{X}$, sigue aproximadamente la distribución $N\left(9000, \frac{866.03}{\sqrt{1000}}\right) = N\left(900, 27.386\right)$. Por tanto, la probabilidad de que una muestra de 1000 elementos tenga un valor medio mayor que 9500 es:

   $P(\overline{X} > 9050) = P\left(\frac{\overline{X}-9000}{27.386} > \frac{9050-9000}{27.386}\right) \approx P\left(Z > 1.825\,8\right) = 0.034$

Ejercicio 61 *En la placa correspondiente a las características de un ascensor se lee: "nº máximo de personas: 6, peso máximo admitido: 450 kg.". La población de usuarios tiene un peso que se distribuye según una Ley Normal de media 60 Kg y desviación típica 20 Kg.*

1. *Hallar la probabilidad de que al montarse 6 personas en el ascensor se supere el peso máximo señalado.*
2. *La indicación de la placa tiene un margen de seguridad. El peligro es real si se superan los 550 kg. ¿A partir de qué número de personas habrá una probabilidad mayor del 10% de que haya peligro real?¿Cual es la probabilidad de peligro real si se suben 6 personas?*

1. Según el Teorema. Central del Límite: "Si $X \in N(\mu, \sigma)$ entonces la suma de los elementos de las muestras de n elementos se distribuyen como una distribución $N\left(n\mu, \sigma\sqrt{n}\right)$. Por lo tanto la suma de los pesos de seis personas se distribuye como una normal de media 360 y desviación típica $20\sqrt{6} = 48.\,990$

   Por tanto la probabilidad de que 6 personas rebasen los los 450 kilos se calcula tipificando la variable:

   $P(Suma > 450) = P\left(\frac{suma-360}{48.990} > \frac{450-360}{48.990}\right) = P\left(\frac{suma-360}{48.990} > 1.\,837\,12\right)$ $= 1 - 0.966\,9 = 3.\,309\,6 \times 10^{-2}$. Al subirse 6 personas en el ascensor la probabilidad de superar el peso máximo es aproximadamente del 3.3%. Que no es poco.

2. $P(Suma > 550) = P\left(Z > \frac{550-60n}{48.990}\right) > 0.10$.Por lo tanto $P\left(Z \leq \frac{550-60n}{48.990}\right) \leq 0.90$; $F^{-1}(0.90) = 1.\,281\,6$; $\frac{550-60n}{48.990} < 1.2816, \quad n > 8.\,120\,2$.Así que a partir de 9 personas la probabilidad de peligro real es mayor del 10%. Luego es bastante peligroso que el número de personas que se suban simultaneamente al ascensor sean 9 o más de 9.

¿Que pasa si se suben 6?
$P\left(\frac{suma-60\times 6}{48.990} > \frac{550-360}{48.990}\right) = P\left(Z > 3.\,878\,3\right) = 1 - 0.999\,95 = 5 \times 10^{-5}$.
Luego si se suben 6 personas la probabilidad de peligro real es prácticamente nula.

T. 5

Inferencia Estadística

Ejercicio 62 *La media de una muestra de 36 elementos de una distribución normal es 4.1. Hallar un intervalo de confianza al 95% para la media. (La desviación típica de la población es 3).*

$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(4.1 - 1.96\frac{3}{\sqrt{36}} < \mu < 4.1 + 1.96\frac{3}{\sqrt{36}}\right)$
$= (4.1 - 0.98, 4.1 + 0.98) = (3.12, 5.08)$.

El valor de z se busca en una tabla de una N(0,1). El valor de la probabilidad que hay que buscar en la tabla es 0.975:

$$P\left(z < z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = 0.975 \Longrightarrow z_{\frac{\alpha}{2}} = 1.96$$

Ejercicio 63 *Se ha repetido un experimento físico 9 veces obteniendose una media de los valores medidos de 42.319 y una cuasi-desviación típica de 5.0 Estimar el valor real de la magnitud con una confianza del 95 por 100*

$\left(\bar{x} - t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(42.319 - 2.31\frac{5}{\sqrt{9}} < \mu < 42.319 + 2.31\frac{5}{\sqrt{9}}\right)$
=(38.469,46.169)

El valor de $t_{\frac{\alpha}{2}}$ =2.31 es el que corresponde a una probabilidad de 0.975 en una tabla de la distribución t de Student con $n - 1 = 8$ grados de libertad

Ejercicio 64 *Para probar si una moneda es defectuosa (la cara y la cruz no tienen la misma probabilidad) se recurre al siguiente ensayo. Se tira la moneda 100 veces y se declara defectuosa si el número de caras es un número fuera del intervalo [40, 60].*

1. *Calcular la probabilidad de declarar la moneda como defectuosa una moneda correcta (error tipo I del test de hipótesis)*

2. *Calcular la probabilidad de declararla correcta si la probabilidad de sacar cara fuera: a) 0.6, b) 0.65, c) 0.70, d) 0.80.*

1. Hay que calcular la probabilidad de que con una moneda correcta ($P(cara) = P(cruz) = 0.5$) se obtenga un número de caras fuera del intervalo [40, 60], cuando se arroja 100 veces. Usamos la aproximación de la binomial $B(100,\ 0.5)$ por la normal $N(100 \times 50, \sqrt{100 \times 0.5 \times 0.5}) = N\,(50, 5)$

   $\alpha = 1 - P(40 \leq x \leq 60) = 1 - P(39.5 \leq x \leq 60.5) =$

   $1 - P(\frac{39.5-50}{5} \leq z \leq \frac{60.5-50}{5}) \approx 1 - P(-2.\,1 \leq z \leq 2.\,1) =$

   $= 1 - (F(2.1) - F(-2.1)) = 3.\,572\,88 \times 10^{-2} \approx 0.036$

2.

$p = prob.$	$\mu = np$	$\sigma = \sqrt{npq}$	$P(40 \leq x \leq 60)$
0.6	60	4.9	$P(\frac{39.5-60}{4.9} \leq z \leq \frac{60.5-60}{4.9}) = 0.46$
0.65	65	4.76	$P(\frac{39.5-65}{4.76} \leq z \leq \frac{60.5-65}{4.76}) = 0.17$
0.70	70	4.6	$P(\frac{39.5-70}{4.6} \leq z \leq \frac{60.5-70}{4.6}) = 0.02$
0.80	80	4	$P(\frac{39.5-80}{4} \leq z \leq \frac{60.5-80}{4}) \approx 0$

   En la tabla anterior se aprecia que conforme la moneda se aparta más de los parámetros correctos va siendo más difícil clasificarla erróneamente como correcta.

Ejercicio 65 *Diseñar una prueba de hipótesis (al 95% de confianza) para la longitud media de una serie de tornillos basada en muestras de 9 elementos, que permita rechazar los lotes cuya longitud media no sea 5 mm. La longitud de estos tornillos se distribuye según una normal de desviación típica $\sigma =$ 2 mm.*

El intervalo de confianza para la media de la muestra al 95% es $\left(\mu - 1.96 \times \frac{\sigma}{\sqrt{n}},\ \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = \left(5 - 1.96 \times \frac{2}{\sqrt{9}},\ 5 + 1.96 \times \frac{2}{\sqrt{9}}\right) =$ $(3.\,693\,33, 6.\,306\,67)$.

Se aceptará que la media es 5 mm. si la media de una muestra de 9 de estos tornillos esta en el intervalo anterior. (Usando este criterio se rechazará injustamente un 5% de lotes cuya media sea 5)

Ejercicio 66 *Un vendedor de bandas elasticas afirma que resisten un estiramiento promedio de 180Kg. Se ha hecho una prueba con 5 de estas banda observandose una resistencia promedio de 169.51Kg. con una cuasi desviación de 5.7 kg*

1. *¿Se rechazaría al 99% de confianza la media de resistencia indicada por el vendedor.*

2. *¿Cual es la región de rechazo para la resistencia promedio de la muestra? ¿Y el valor Crítico?*

1. El estadístico de contraste es $T = \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}$ que se distribuye como una t de Student con n-1 grados de libertad. En este caso vamos a usar como hipótesis alternativa que el estiramiento sea menor que 180 kg., ya que no nos parece dañino que las bandas tengan más resistencia que la que declara el fabricante.

   Rechazamos la afirmación del vendedor si $T < t_{4,\, 0.01}$; $P\left(T < t_{4,\, 0.01}\right) = 0.01 \Rightarrow t_{4,\, 0.01} = T_4^{-1}(0.01) = -3.746\,95$

   En este caso es $T = \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{169.51-180}{\frac{5.7}{\sqrt{5}}} = -4.115\,15 < -3.746\,95 = t_{4,\, 0.01}$

   Por lo tanto rechazamos la declaración del fabricante.

2. El valor crítico es el que hace $T = \frac{\overline{X}-180}{\frac{5.7}{\sqrt{5}}} = t_{4,\, 0.01} = -3.746\,95$,que resulta ser $170.\,449$. La región de rechazo es $(-\infty, 170.449)$. Es decir que si la resistencia media de la muestra de 5 elementos es menor que 170.449Kg. rechazamos la media del fabricante, inclinandonos por la opción de que la resistencia media sea menor que 180 kg.

Ejercicio 67 *Un tipo de botes de pintura esta declarada como apta para pintar un promedio de 80 m^2 con una desviación típica de 8.4 m^2.Se desea comprobar si puede aceptarse este valor promedio. Con este objetivo se ha decidido probar 100 de estos botes y rechazar la pintura si el promedio de superficie pintada resultará menor que 78 m^2 Se aceptará el valor de la desviación típica.*

1. *Calcular el nivel de confianza y la significación de esta prueba.*

2. *Si la pintura pintara ralmente un promedio de 79 m^2 cual sería la probabilidad de no rechazar la media indicada por el fabricante.*

3. *¿Y si el promedio fuera de 75 m^2*

1. Realizamos una prueba unilateral. Si μ fuera realmente 80

   $P\left(\overline{x} < 78\right) = P\left(\frac{\overline{x}-80}{\frac{8.4}{\sqrt{100}}} < \frac{78-80}{\frac{8.4}{\sqrt{100}}} = -2.3809\right) = 0.0087$, sería la probabilidad de cometer un error tipo I $\Rightarrow$ $\alpha = 0.0087$

   El nivel de confianza sería entonces 99.13%

2. El error tipo II (probabilidad de aceptar la hipótesis nula siendo falsa) sería
   $P(\overline{x} > 78) = P\left(\frac{\overline{x}-79}{\frac{8.4}{\sqrt{100}}} > \frac{78-79}{\frac{8.4}{\sqrt{100}}} = -1.190\,48\right) = 0.88$

3. Realizando los mismos cálculos sustituyendo 79 por 75 resulta en este caso
   $P(\overline{x} > 78) = P\left(\frac{\overline{x}-75}{\frac{8.4}{\sqrt{100}}} > \frac{78-75}{\frac{8.4}{\sqrt{100}}} = 3.571\,43\right) = 0.00017$

   Por lo tanto es facil aceptar erroneamente una pintura que pinta por promedio de 79 m^2, pero dificil aceptarla si este promedio fuera de 75 m^2

Ejercicio 68 *Un vendedor de neumáticos dice que la vida media de sus neumáticos es de 28000 Km. Admitiendo para la desviación típica el valor 1348 Km. diseñar un test de hipótesis al 99% de confianza, basado en muestras de 40 elementos que permita contrastar la hipótesis nula de ser* $\mu = 28000Km$ *usando como hípótesis alternativa* $\mu < 28000Km$

$P(\overline{x} < c) = P\left(\frac{\overline{x}-28000}{\frac{1348}{\sqrt{40}}} < \frac{c-28000}{\frac{1348}{\sqrt{40}}}\right) = 0.01 \Rightarrow \frac{c-28000}{\frac{1348}{\sqrt{40}}} = -2.33$ y por tanto $c = 27503.4$

La prueba consiste en ensayar 40 neumáticos. Aceptariamos $\mu = 28000Km$ si el promedio de vida de 40 neumáticos es al menos 27503.4 Km. Si el promedio de duración fuese menor que 27503.4 Km nos inclinaríamos por la opción $\mu < 28000Km$.

Ejercicio 69 *Si de un total de 100 personas entrevistadas 36 han afirmado que conocen una cierta marca de detergente*

1. *Hallar un intervalo de confianza al 95% para la proporción real de personas que conocen este detergente.*

2. *¿Cuantas personas se precisan entrevistar para que el intervalo de confianza para la proporción tenga una amplitud de 0.1?*

1. Si la variable es número de personas x de cada 100 que conocen el detergente usamos el modelo $x \in B(n,p)$. La aproximación normal de la binomial es $N\left(np, \sqrt{npq}\right)$. La proporción de personas de cada muestra que conoce el detergente es $\frac{x}{n}$, que se distribuye con una $\mathrm{N}\left(p, \sqrt{\frac{pq}{n}}\right)$. El intervalo de confianza sería:

$$\left(p - 1.96\sqrt{\frac{pq}{n}}, p - 1.96\sqrt{\frac{pq}{n}}\right)$$

Tomamos como valor central la estimación muestral de p y al valor máximo para amplitud del intervalo Así que sería:

$$\left(0.36 - 1.96\sqrt{\frac{0.5 \times 0.5}{100}},\ 0.36 + 1.96\sqrt{\frac{0.5 \times 0.5}{100}}\right) =$$

$$= \left(0.36 - 1.96\sqrt{\frac{0.5 \times 0.5}{100}},\ 0.36 + 1.96\sqrt{\frac{0.5 \times 0.5}{100}}\right) = (0.262, 0.458)$$

el intervalo de confianza para la proporción poblacional.

2. El radio del intervalo ha de ser $1.96\sqrt{\frac{0.5\times0.5}{n}} < \frac{0.1}{2} = 0.05$

   resolviendo la ecuación $1.96\sqrt{\frac{0.5\times0.5}{n}} = 0.05$ resulta para n el valor 384. 16. Por tanto el número de personas entrevistadas ha de ser al menos 385.

Ejercicio 70 *Se desea saber la proporción de personas de una gran ciudad que encuentran adecuado el transporte público. ¿ Cuántas personas hay que entrevistar si se desea estimar esta proporción con un intervalo de confianza de 95% y un error de precisión menor del 6%?.*

$z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \leq 0.06 \Rightarrow 1.96\sqrt{\frac{pq}{n}} \leq 0.06$.Tomando el mayor valor para el radio del intervalo ($p = 0.5,\ q = 0.5$) resulta, $1.96\sqrt{\frac{0.25}{n}} \leq 0.06$ y $n \geq 267$. Así que hay que entrevistar a 267 personas.

Ejercicio 71 *Encuestadas 267 personas ha resultado que 114 de ellas encuentran satisfactorio el transporte público. Dar un intervalo de confianza para la proporción de personas que encuentran satisfactorio este tipo de transporte.(95% de confianza)*

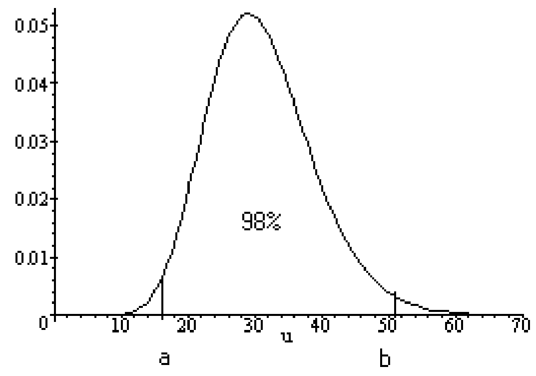
$$\left(\frac{114}{267} - 1.96\sqrt{\frac{0.25}{267}}, \frac{114}{267} + 1.96\sqrt{\frac{0.25}{267}}\right) = (0.366\,99,\ 0.486\,94)$$

Ejercicio 72 *32 medidas del punto de ebullución del azufre tienen una cuasidesviación de 0.83 grados. Calcular un intervalo de confianza para la varianza con una confianza del 98%*

Usamos el hecho demostrado de que

$$\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$$

esto es, una chi-cuadrado con $n - 1$ grados de libertad.En este caso sería tendría 31 grados de libertad.



$P(x < \text{a}) = 0.01 \Rightarrow a = 15.655$

$P(x > \text{b}) = 0.01 \Rightarrow b = 52.119$

$\left(15.655 < \frac{31s^2}{\sigma^2} < 52.119\right)$. Tomando $\text{s}^2 = 0.83^2 = 0.688\,9$. se deduce un intervalo de confianza para σ^2, $\quad 0.4091 < \sigma^2 < 1.3641$.

Ejercicio 73 *Las piezas de una maquina deben ser del mismo tamaño, por eso se exige que la desviación típica de la población sea 0.05 mm. Diseñar un test al 95% de confianza para contrastar la hipótesis de que* $\sigma = 0.05$ *mm. con muestras de 15 elementos*

$\frac{14s^2}{\sigma^2}$ es una chicuadrado con 14 grados de libertad

$P\left(\chi^2_{14} > 23.685\right) = 0.05$

$\frac{14s^2}{0.05^2} = 23.685 \Rightarrow s^2 = \frac{23.685\times(0.05)^2}{14} = 4.\,229\,46 \times 10^{-3}$.

En consecuencia si la desviación típica de la muestra de 15 elementos fuera mayor que $\sqrt{4.\,229\,46 \times 10^{-3}} = 6.\,503\,43 \times 10^{-2} = 0.065$, rechazamos el valor 0.05 para la desviación típica de la población, concluyendo que es muy posible que sea mayor.

Ejercicio 74 *Se ha llevado a cabo un estudio para determinar si hay diferencia entre el tiempo que tardan los hombres y las mujeres en hacer determinada maniobra en una línea de ensamble. Los valores obtenidos en el estudio se resumen en la siguiente tabla*

	N° de elementos	*media muestral*	*Varianza poblacional*
hombres	*50*	*42 seg.*	*18 seg*^{2}
mujeres	*50*	*38 seg*	*14 seg*^{2}

¿Es significativa la diferencia de rendimiento entre hombres y mujeres?
$(\overline{x}_1 - \overline{x}_2) \in N\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_1}}\right)$

La hipótesis nula es $\mu_1 - \mu_2 = 0$, la hipótesis alternativa es $\mu_1 - \mu_2 \neq 0$
$z = \frac{42-38-0}{\sqrt{\frac{18}{50}+\frac{14}{50}}} = 5.0$. El valor crítico al 95% de confianza es $1.96 << 5$, lo que parece indicar que el mejor valor de la media masculina es real, y por tanto la diferencia de rendimiento es signifícativa.

Ejercicio 75 *Un fabricante asegura que sus fusibles, con una sobrecarga del 20%, se fundiran por promedio al cabo de 12.40 min. Una muestra de 20 fusibles se sobrecarga un 20%, obteniendose una media de 10.63 y una cuasidesviación de 2.48 min. ¿Confirma la muestra la afirmación del fabricante para el promedio?*

Como la muestra es pequeña y la varianza desconocida, hay que estimarla con la muestra. Por eso hay que usar la t de Student en el test de hipótesis:
$t = \frac{10.63-12.40}{\frac{2.48}{\sqrt{20}}} = -3.19181$
El valor crítico para la t de Student con 19 grados (95% de confianza) es
$T_{19}^{-1}(0.025) = -2.09302$.
El valor experimental es más pequeño que -2.09302, luego se rechaza la hipótesis nula. No se confirma la afirmación del fabricante.

Ejercicio 76 *Se han recogido muestras de aire para estudiar su contaminación, obteniendose las siguientes cantidades de impurezas en* $\frac{Kg}{m^3}$

$$2.2;\ 1.8;\ 3.1;\ 2.0;\ 2.4;\ 2.0;\ 2.1;\ 1.2$$

Dad un intervalo de confianza al 95% para la media de impurezas contenidas en el aire

Calculamos la media y la cuasi desviación de los valores de la muestra, que resultan:

$$\overline{x} = 2.1, \ s = 0.537$$

Ya que la muestra tiene solo 8 elementos usamos para calcular el intervalo de confianza el valor correspondiente a la t de Student con 7 grados de libertad

$$\left(2.1 - 2.364 \times \frac{0.537}{\sqrt{8}}, \ 2.1 + 2.364 \times \frac{0.537}{\sqrt{8}}\right) = (1.651\,18, 2.548\,82)$$

Ejercicio 77 *El director de un colegio quiere saber el tiempo medio que tardan los alumnos en cambiar de clase, con una confianza del 99% y un error que no sobrepase 0.25 mininutos. Si se puede suponer que el valor de σ es 1.40 minutos, ¿Cuál debe ser el tamaño de la muestra?*

Usando la distribución normal, ya que se supone conocida la desviación típica de la población

$$z_{0.995} \times \frac{1.4}{\sqrt{n}} \leq 0.25; \qquad 2.5758 \times \frac{1.4}{\sqrt{n}} \leq 0.25$$

de aquí resulta que

$$n \geq 208.06$$

Tomaremos una muestra de 209 alumnos

Ejercicio 78 *Se realizó un muestreo para decidir si los sueldos de los peones de albañil de una ciudad A y de otra B son iguales por promedio o no. Para ello se consulto a 100 peones de la ciudad A y a 150 de la ciudad B. Analizadas la respuestas realizadas por dichos operarios se determino que la media de los sueldos de los 100 operarios de la ciudad A era de 760 € y la de los 150 empleados de ciudad B era de 720 €. Suponiendo que la desviación típica poblacional de los sueldos de A es 12€ y la de B 9 €, decidir si el sueldo medio en ambas ciudades es igual o distinto.*

Las muestras son independientes y las varianzas conocidas. El estadístico de contraste es

$$Z = \frac{\left(\overline{X_1} - \overline{X_2}\right) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

que se distribuye como una normal estándar. En este caso su valor es

$$Z = \frac{\left(\overline{X_1} - \overline{X_2}\right) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(760 - 720) - 0}{\sqrt{\frac{12^2}{100} + \frac{9^2}{150}}} = 28.\,427$$

Considerando un intervalo de confianza del 95% el intervalo de aceptación es $[-1.96, 1.96]$. Por tanto el valor experimental queda claramente fuera de este intervalo, así que se rechaza la hipótesis de igualdad entre las medias de los sueldos de estos empleados. por tanto la diferencia hallada entre las medias es significativa.

Ejercicio 79 *Se desea comparar el gasto medio mensual en alimentación entre las familias de dos barrios. Para ello se seleccionaron 20 familias de cada barrio, observando sus gastos mensuales en alimentación. Se determino la media y las cuasidesviaciones típicas, obteniéndose los siguientes resultados muestrales:* $\left(\overline{X_1} = 200,\ S_1 = 20 \qquad \overline{X_2} = 175,\ S_2 = 17\right)$. *Suponiendo que los gastos se distribuyen normalmente decidir sobre la cuestión planteada. Los gastos medios en alimentación entre ambos barrios, ¿pueden considerarse iguales?*

Contrastamos en primer lugar la igualdad entre las varianzas. considerando muestras independientes. El estadístico de contraste es $F = \frac{S_1^2}{S_2^2}$, que se distribuye como una F_{n_1-1,n_2-1}.

En el caso del ejercicio sería $F = \frac{20^2}{17^2} = 1.\,384\,1$. El intervalo de aceptación al 95% para F es $[F_{19,19}^{-1}(0.025), F_{19,19}^{-1}(0.975] = [0.395\,81, 2.\,526\,5]$. Por tanto considero que las varianzas son iguales.

Se realiza ahora el test para contrastar la igualdad entre las medias con dos muestras independientes, en el caso de que las varianzas se consideren iguales. El estadístico de contraste es:

$$T = \frac{\left(\overline{X_1} - \overline{X_2}\right) - 0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, siendo S = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1 + n_2 - 2}}$$

Ahora $S = \sqrt{\frac{20^2(20-1)+17^2(20-1)}{20+20-2}} = 18.\,561 = 18.\,321$ y por tanto $T = \frac{(200-175)-0}{18.561\sqrt{\frac{1}{20}+\frac{1}{20}}} = 4.259\,3$.

la región de aceptación para la t de Student con 38 grados de libertad al 95% de confianza es:

$$[T_{38}^{-1}(0.025), T_{38}^{-1}(0.975] = [-2.\,024\,4, 2.\,024\,4]$$

Por tanto, la decisión sugerida por el test es rechazar la igualdad entre los valores medios de gastos entre ambos barrios, puesto que 4.259 3 no pertenece a este intervalo.

Ejercicio 80 *Mendel sembró 532 plantas de guisantes usando semillas del mismo tipo y los frutos resultantes los clasificó atendiendo al color en: verde, verde amarillento y amarillo y atendiendo a la forma: redondo, levemente rugoso y rugoso. Obtuvo los siguientes datos:*

	Verde	*Verde-Amarillo*	*Amarillo*	
Redondo	*35*	*68*	*38*	*141*
Levemente Rugoso	*67*	*138*	*60*	*265*
Rugoso	*30*	*68*	*28*	*126*
	132	*274*	*126*	*532*

¿Había alguna relación de dependencia entre la forma y el color de esos guisantes?

Emplearemos el test Chi-Cuadrado apropiado para tablas de contingencia. Calculamos en primer lugar las frecuencias esperadas en cada casilla si las variables color y forma fueran independientes:

Por ejemplo, en la primera casilla, deben estar los guisantes verdes y redondos: Si suponemos que ambas cualidades, forma y color son independientes, entonces $P(V \cap R_e) = P(V)P(R_e) = \frac{132}{532} \times \frac{141}{532} = 6.576\,1 \times 10^{-2}$

Por tanto , bajo la hipótesis de independencia, el número esperado de guisantes en las primera casilla es $np_1 = 532 \times 6.58 \times 10^{-2} = 35.006$

Mostramos ahora los cálculos que habría que realizar en la siguiente casilla de la derecha:

$np_2 = 532 \times P(VA \cap R_e) = 532 \times \frac{274}{532} \times \frac{141}{532} = 72.62.$

De manera similar se rellenan el resto de las casillas obteniéndose la siguiente tabla de valores esperados

	Verde	Verde-Amarillo	Amarillo	
Redondo	35.006	72.62	33´4096	141
Levemente Rugoso	65´7552	136´458	62´77667	265
Rugoso	31´2816	64´90468	29´845230	126
	132	274	126	532

Calculamos el valor de la Chi-cuadrado experimental:

$$\chi^2_{\exp} = \sum_{i=1}^{9} \frac{(n_i - np_i)^2}{np_i}$$

realizando las operaciones indicadas obtenemos:

$$\chi^2_{\text{exp}} = \sum_{i=1}^{9} \frac{(n_i - np_i)^2}{np_i} = \frac{(35-35.006)^2}{35.006} + \frac{(68-72.618)^2}{72.618} + \cdots + \frac{(68-64.904)^2}{64.904} + \frac{(28-29.8452)^2}{29.8452} = 1.4024$$

este valor debe compararse con el valor teórico que corresponde a $\chi^2_{(c-1)(f-1)} = \chi^2_{2\times 2} = \chi^2_4$ al nivel de significación que se requiera.

Empleando en esta ocasión un nivel de significación del 95% obtenemos $\left(\chi^2_4\right)^{-1}(0.95) = 9.487\,7$, que es el mayor valor aceptable para χ^2_{exp}. Como el valor obtenido, 1.4024, es bastante menor que 9.487 7, consideramos que puede aceptarse la hipótesis de independencia entre las características de forma y color de los guisantes, al 95% de confianza.

Unidad Temática II

PROBLEMAS DE CONTROL DE CALIDAD

T. 6

Introducción. Control de Atributos

Ejercicio 81 *Se tomaron 25 muestras con 100 lamparas cada una, conteniendo los siguientes números de defectos:*

3, 4, 6, 4, 0, 5, 2, 3, 0, 2, 3, 5, 3, 9, 1, 2, 4, 4, 1, 8, 4, 6, 5, 3, 2

1. *¿Se puede aceptar que este proceso produce un 2% de defectuosos por promedio como afirma el fabricante?*

2. *Con los datos anteriores calcular unos nuevos límites de control para la fracción de lamparas defectuosas*

3. *¿Cual debe ser la proporción media de defectos que debe dar el fabricante?*

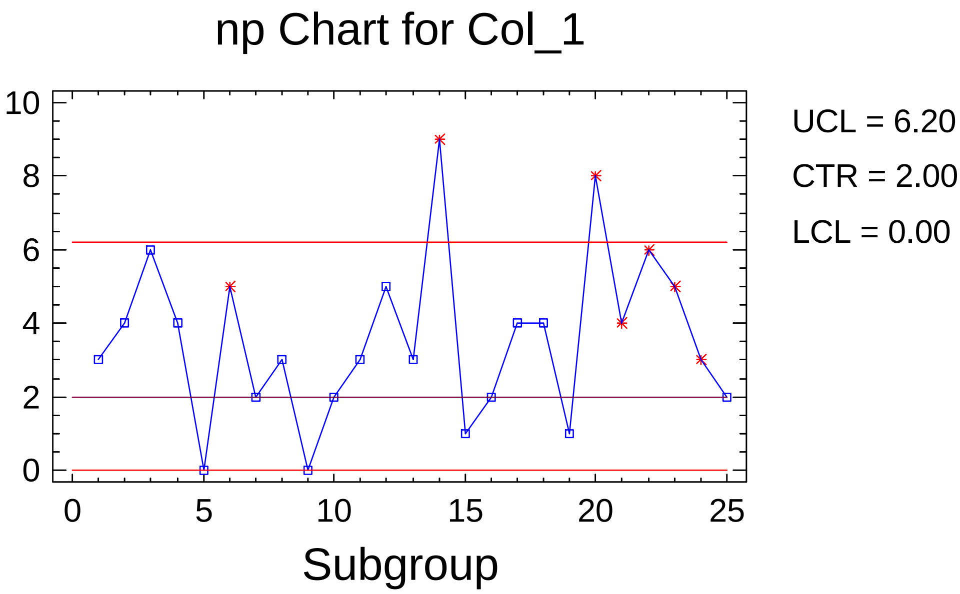
1. Si el número promedio de defectos fuera del 2%, sería $p = 0.02$.

   $\left(p - 3\sqrt{\frac{p(1-p)}{100}}, p + 3\sqrt{\frac{p(1-p)}{100}}\right) =$

   $=\left(0.02 - 3\sqrt{\frac{0.02\times0.98}{100}}, 0.02 + 3\sqrt{\frac{0.02\times0.98}{100}}\right) =$

   $(0.02 - 0.042\,, 0.02 + 0.042\,) = (-.0\,022, .0\,062)$.

   Los límites de control para p serían $(0, 0.062)$. Así que los límites de control para el número de defectos en muestras de 100 elementos serían serían. (0,6.2). El proceso no está bajo control, ya que hay dos muestras fuera de control (los valores 8 y 9). La situación puede verse reflejada en el siguiente gráfico de control.



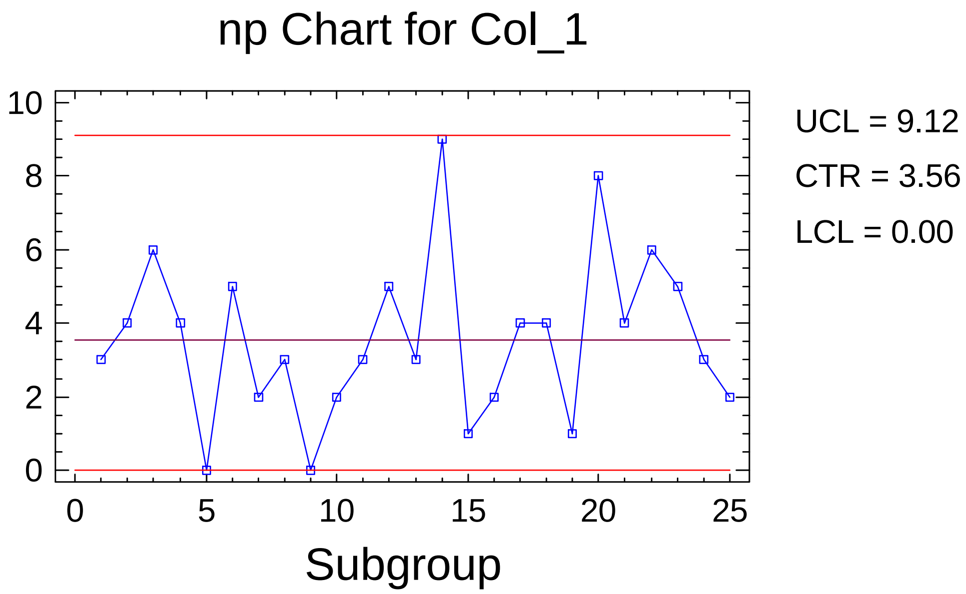
2. Utilizando los datos de la muestra obtenemos una media para el número de defectos 3.56 y para la proporción de defectuosos $p = 0.0356$. Los límites de control para la proporción de defectos son ahora:

$$\left(p - 3\sqrt{\frac{p(1-p)}{100}}, p + 3\sqrt{\frac{p(1-p)}{100}}\right) =$$

$$= \left(0.0356 - 3\sqrt{\frac{0.0356 \times 0.9644}{100}}, 0.0356 + 3\sqrt{\frac{0.0356 \times 0.9644}{100}}\right) =$$

$$= (-0.019987, \ 0.091187)$$

Así que los límites de control para la proporción sería (0, 0.091187) y para el número de defectos con muestras de 100 elementos (0, 9.1187). Ahora no hay ningún número fuera de estos límites, así que el proceso está bajo control. La representación de estos nuevos límites de control y del número de defectos de las 25 muestras se observa en el siguiente Diagrama de control:



$\widehat{p} = 0.0356$, 3.56% se debería dar como promedio de n^o de lamparas defectuosas produce este proceso

Ejercicio 82 *Durante la fabricación de piezas de un aparato eléctrico se han tomado muestras de 50 elementos cada 4 horas. Se han registrado la cantidad de elementos defectuosos entre estas 50 piezas: 3, 3, 2, 0, 6, 1, 1, 1, 2, 1, 2, 3, 3, 0, 8, 0, 6, 5, 5, 0, 3, 3, 2, 1, 3, 4, 5, 3, 4, 5, 4, 6, 1, Utilizar estos datos para establecer unos valores de control para la proporción de elementos defectuosos que produce este proceso.*

La media de la muestra, número medio de defectos es $\frac{96}{33} = 2.91 = n\widehat{p}$ y la estimación de p a partir de la muestra es $\frac{2.91}{50} = 0.058$

Un posible intervalo de control para el número de elementos defectuosos es

$$\left(n\widehat{p} - 3\sqrt{n\widehat{p}\widehat{q}}, n\widehat{p} + 3\sqrt{n\widehat{p}\widehat{q}}\right) =$$

$$= \left(2.91 - 3\sqrt{2.91 \times 0.942}, 2.91 + 3\sqrt{2.91 \times 0.942}\right) =$$
$$= (-2.057, 7.877)$$

Como el valor 8 de la muestra ha quedado fuera de los límites de control, hay que repetir el cálculo anterior eliminando el valor 8 de la muestra.

La media de la muestra en este segundo caso sería $88/32 = 2.75$ y $p = 2.75/50 = 0.055$

Repitiendo el proceso omitiendo el valor 8 se obtiene los límites de control para el número de defectos:

$$\left(n\widehat{p} - 3\sqrt{n\widehat{p}\widehat{q}}, n\widehat{p} - 3\sqrt{n\widehat{p}\widehat{q}}\right) =$$
$$\left(2.75 - 3\sqrt{2.75 \times 0.945}, 2.75 + 3\sqrt{2.75 \times 0.945}\right) = (-2.086\,2, 7.586)$$

Ahora el intervalo que debe contener el número de defectos de las 32 muestras es (0, 7.58). Como no hay valores fuera de estos límites se aceptan como límites de control para el número de defectos.,Por tanto, el intervalo de control para la proporción de defectos es $(0, \frac{7.58}{50}) = (0,\ 0.151\,6)$

Ejercicio 83 *Los siguientes datos son el número de soldaduras defectuosas encontradas en sucesivas muestras de 500 juntas soldadas:106, 116, 164, 89, 99, 40, 112, 36, 69, 74, 42, 37, 25, 88, 101, 64, 51, 74, 71, 43, 80. ¿Está el proceso bajo control.?*

Una estimación para p se calcula hallando la media de los valores 106, 116, 164, 89, 99, 40, 112, 36, 69, 74, 42, 37, 25, 88, 101, 64, 51, 74, 71, 43, 80. La media resulta $75.\,286$.Dividiendo este valor por el número de elementos de cada muestra (500) se obtiene una estimación para media de la proporción de defectos que resulta 0 .150 57.

Los límites de control para la proporción de defectuosos son:

$$(\widehat{p} - 3\sqrt{\frac{\widehat{p}\widehat{q}}{n}}, \widehat{p} + 3\sqrt{\frac{\widehat{p}\widehat{q}}{n}}) = (0.15 - 3\sqrt{\frac{0.15 \times 0.85}{500}}, 0.15 + 3\sqrt{\frac{0.15 \times 0.85}{500}}) =$$
$$(0.102\,09, 0.197\,91)$$

El número de defectuosos se obtiene multiplicando por el número de elementos de la muestra, 500 en este caso. Por tanto $(0.102\,09 \times 500, 0.197\,91 \times 500) = (51.\,045, 98.\,955)$ son los límites de control propuestos.

Como hay demasiados elementos fuera de estos límites, no parece que el proceso esté bajo control.

Ejercicio 84 *En una planta industrial se encapsulan las botellas de una bebida refrescante. Cada 5 horas se seleccionan 64 de estas botellas para comprobar si la operación se ha realizado correctamente, resultando que la proporción media de botellas defectuosas ha sido del 2 %.*

1. *Suponiendo que no hay muestras fuera de control, definir los límites de control para el número de botellas defectuosas de cada muestra de 64 botellas.*

2. *Si una persona compra 12 de estas botellas ¿Cúal es la probabilidad de que ninguna de ellas sea defectuosa?*

3. *Un cliente de esta fabrica, que no sabe cuantas defectuosas se producen decide que no las adquirirá si en una muestra de 100 de ellas el número de defectuosas es mayor que 1. ¿Cual es la ordenada, para p = 0.05, de la curva característicacorrespondiente.*

4. *Además este cliente decide que si durante la siguiente semana (7 dias) hay más de dos días en que no puede adquirir las botellas dejará de confiar en este prooveedor y ya no le comprará más. Calcular la probabilidad que tiene la fábrica de perder este cliente.*

1. $\left(np - 3\sqrt{npq}, np + 3\sqrt{npq}\right) =$
   $=\left(64 \times 0.02 - 3\sqrt{64 \times 0.02 \times 0.98}, 64 \times 0.02 + 3\sqrt{64 \times 0.02 \times 0.98}\right) =$
   $= (-2.\,08, 4.\,64)$. Tomamos como limites $(0.4.64)$

2. $\binom{12}{0} 0.02^0 \times 0.98^{12} = 0.784\,72.$
   Usando la aproximación normal
   $N(np, \sqrt{npq}) = N(12 \times 0.02, \sqrt{12 \times 0.02 \times 0.98}) = N(0.24, 0.484\,97)$
   $P(x < 0.5) = 0.704\,06$

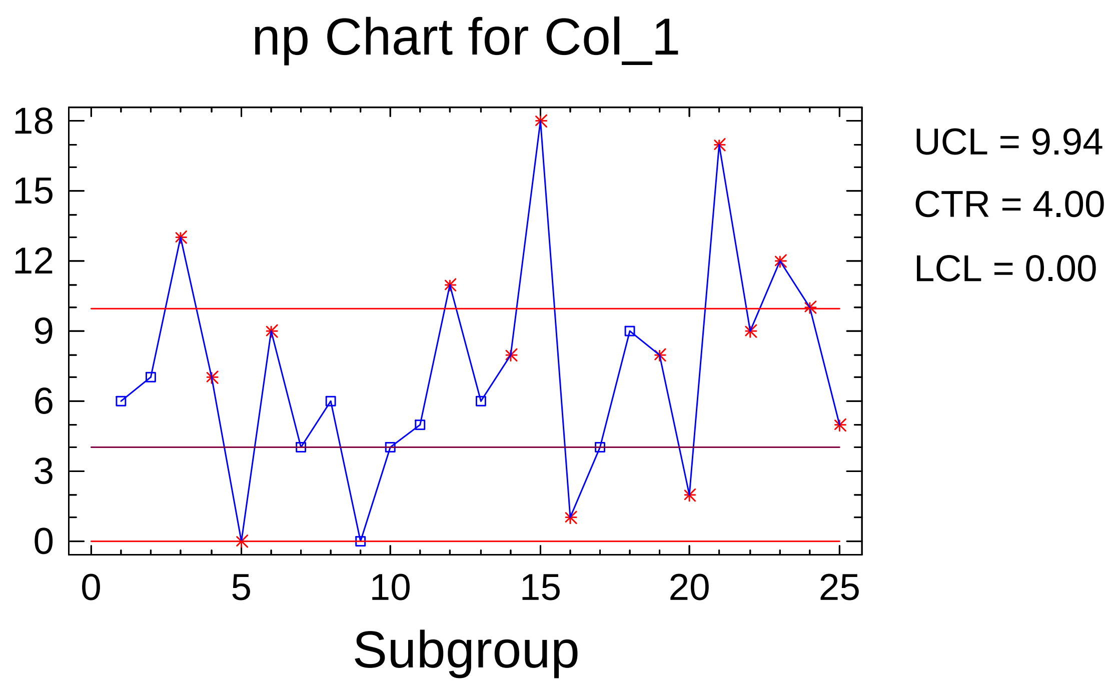
3. P(n^o de defectos ≤ 1) $= \binom{100}{0} 0.05^0 \times 0.95^{100} + \binom{100}{1} 0.05^1 \times 0.95^{99} = 3.\,708\,1 \times 10^{-2}$
   Usando la aproximación normal
   $N(np, \sqrt{npq}) = N(100 \times 0.05, \sqrt{100 \times 0.05 \times 0.95}) = N\,(5.0, 2.\,179\,4)$
   P(n^o de defectos ≤ 1) $= P(x \leq 1.5) = 5.\,414\,3 \times 10^{-2}$

4. Probabilidad de adquirirla un cierto día
   $\binom{100}{0} 0.02^0 \times 0.98^{100} + \binom{100}{1} 0.02^1 \times 0.98^{99} = 0.403$
   Usando la aproximación por la $N(2,\ 1.4)$
   P(n^o de defectos $\leq\ 1$) $= P(x \leq 1.5) = 0.\,360$
   Probabilidad de perder el cliente$= \sum_{i=3}^{7} \binom{7}{i} \left(1 - 0.\,403\,27\right)^i \times .\,403\,27^{7-i} =$ 0.900

Ejercicio 85 *25 muestras sucesivas de 200 interruptores tomadas de una línea de producción contuvieron respectivamente 6, 7, 13, 7, 0, 9, 4, 6, 0, 4, 5, 11, 6, 8, 18, 1, 4, 9, 8, 2, 17, 9, 12, 10, 5 piezas defectuosas. Se quiere mantener la fracción de piezas defectuosas en 0.02. Elabore un diagrama con estos datos para indicar si se cumple esta norma o no.*

Los límites de control que se han de mantener son
$(200 \times 0.02 - 3\sqrt{200 \times 0.02 \times 0.98},\ 200 \times 0.02 + 3\sqrt{200 \times 0.02 \times 0.98}) = (-1.9397, 9.9397) = (0,\ 9.93937)$.

En la muestra hay 6 elementos que superan el valor 9.3937, así que no se puede afirmar que se cumple la norma ya que hay 6 muestras fuera de control. El diagrama correspondiente a la comparación de los datos con la norma aparece en la siguiente figura.



Ejercicio 86 *Los siguientes valores*

$$18,\ 15,\ 23,\ 9,\ 27,\ 19,\ 22,\ 21,\ 25,\ 14,$$
$$19,\ 26,\ 11,\ 28,\ 22,\ 14,\ 25,\ 17,\ 23,\ 18$$

son el número de unidades defectuosas en muestras de 200 componentes electrónicos de los producidos en un cierto proceso.

1. *Estimar a partir de estas muestras una estimación para la proporción de componentes electrónicos defectuosos que produce este proceso.*

2. *Calcular los límites de control para esta proporción a partir de estas muestras*

3. *Calcular la capacidad del proceso*

1. 18+15+23+9+27+19+22+21+25+14+19+26+11+28+22+14+25+ $17 + 23 + 18 = 396$
   $\widehat{p} = \frac{396.0}{20 \times 200} = 0.099$

2. Intervalo de confianza para la proporción de defectuosos es
   $\left(p - 3\sqrt{\frac{pq}{n}}, p + 3\sqrt{\frac{pq}{n}}\right) =$
   $\left(0.099 - 3\sqrt{\frac{0.099 \times 0.901}{200}}, 0.099 + 3\sqrt{\frac{0.099 \times 0.901}{200}}\right) = (0.0356,\ 0.1624)$
   El intervalo de confianza para el número de elementos defectuosos es
   $200\left(3.5644 \times 10^{-2}, .16236\right) = (7.1288, 32.472)$, por lo tanto no hay muestras fuera de control

3. La capacidad del proceso es $1 - p = 0.901$

Ejercicio 87 *Se inspeccionan las botellas de plástico para un detergente líquido. Se toman 20 muestras cada una con 100 botellas, notificandose la fracción de defectuosas de cada muestra. Los datos aparecen a continuación*

Muestra	*Fracción de defectuosas*	*Muestra*	*Fracción de defectuosas*
1	*0.12*	*11*	*0.13*
2	*0.15*	*12*	*0.07*
3	*0.18*	*13*	*0.12*
4	*0.10*	*14*	*0.08*
5	*0.12*	*15*	*0.09*
6	*0.11*	*16*	*0.15*
7	*0.05*	*17*	*0.10*
8	*0.09*	*18*	*0.06*
9	*0.13*	*19*	*0.12*
10	*0.10*	*20*	*0.13*

1. *Estimar a partir de estas muestras un valor para la proporción de botellas defectuosa que se fabrican en este proceso*

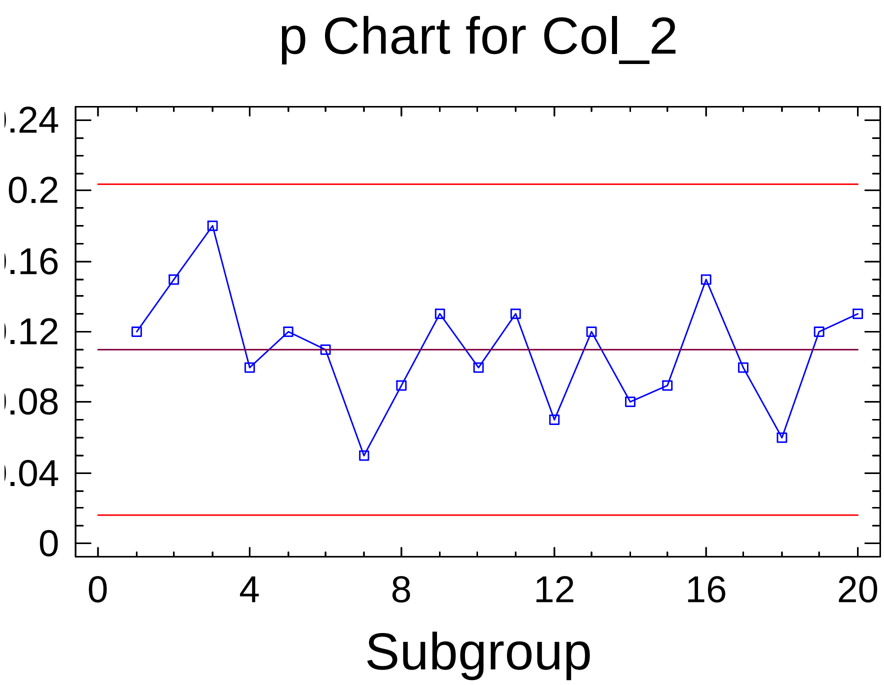
2. *Calcular los límites de control a partir de estas muestras*

3. *Calcular la capacidad del proceso*

1. Realizando la media de las fracciones de defectuosas contenidas en los datos, se obtiene 0.11:por tanto $\widehat{p} = 0.11$

2. Los límites de control para la fracción de defectuosos se obtiene por medio del intervalo de confianza:
   $\left(p - 3\sqrt{\frac{pq}{n}}, p - 3\sqrt{\frac{pq}{n}}\right) = \left(0.11 - 3\sqrt{\frac{0.11 \times 0.89}{100}}, 0.11 + 3\sqrt{\frac{0.11 \times 0.89}{100}}\right) = \left(1.6133 \times 10^{-2}, 0.20387\right) = (0.016, 0.20)$

   Todos los valores muestrales están dentro de estos límites, así que el proceso está bajo control y los valores calculados son los límites superior e inferior de control. La situación esta representada en el siguiente diagrama:



3. La capacidad del proceso es $1 - p = 0.89$

Ejercicio 88 *Dentro de un proyecto de mejora de la calidad, una industria textil decide controlar el número de imperfecciones encontradas en cada pieza de tela. Se estima que el número promedio de imperfecciones por cada pieza de tela es de 12. Calcular la probabilidad de que en una de estas piezas de tela fabricada se encuentren.*

1. *Entre 10 y 12 imperfecciones.*

2. *Menos de 8 y más de 16 imperfecciones.*

3. *Inspeccionada un lote de 25 piezas de tela, se han encontrado los siguientes números de defectos:*

   13, 15, 9, 7, 12, 8, 4, 10, 3, 5, 8, 14, 10, 11, 14, 15, 7, 16, 8, 8, 9, 14, 17, 13 , 9. ¿Se mantiene el número promedio de defectuosos 12? Realizar la gráfica de control

1. El número de defectos por unidad producida suele modelarse con una distribución de Poisson, cuya función de probabilidad es: $P(x=i) = \frac{\lambda^i}{i!}e^{-\lambda}$, siendo en este caso $\lambda = 12$.

   Por tanto la probabilidad pedida se calcularía:

   $P(10 \leq x \leq 12) = \frac{12^{10}}{10!}e^{-12} + \frac{12^{11}}{11!}e^{-12} + \frac{12^{12}}{12!}e^{-12} = 0.333\,57$

   Debido a que este cálculo es a veces dificultoso, se suele aceptar la aproximación con la normal de la misma media y varianza, si $\lambda > 10$.

   Si se realiza el cálculo aproximado empleando la distribución $N(12, \sqrt{12}) = N(12, 3.\,464\,1)$, se calcula $P(9.5 \leq x \leq 12.5)$, ya que es preciso usarla correción por continuidad por paso de distribución discreta a continua. En este caso el resultado obtenido sería.

   $P(9.5 \leq x \leq 12.5) = P(x \leq 12.5) - Px \leq 9.5) = 0.557\,38 - 0.235\,24 = 0.322\,14$

2. Empleando la aproximación normal, la probabilidad pedida es:

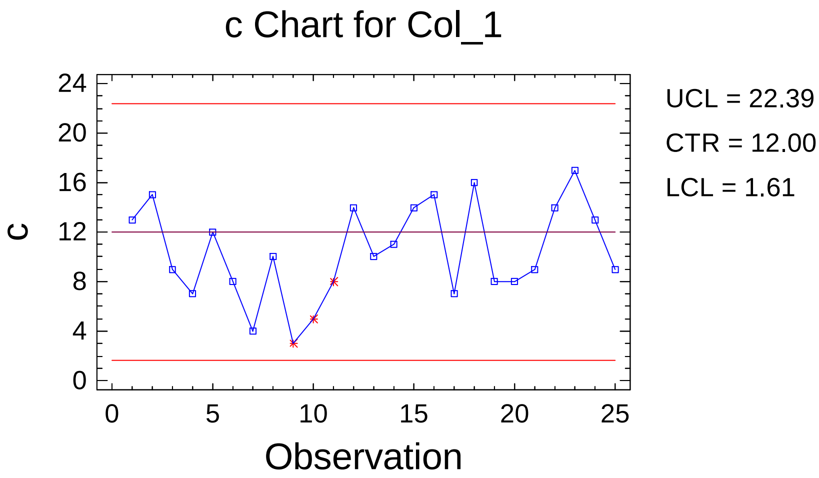
   $1 - P(7.5 \leq x \leq 16.5) = 1 - [P(x \leq 16.5) - P\,(x \leq 7.5))] =$

   $= 1 - \left(0.903\,03 - 9.\,696\,5 \times 10^{-2}\right) = 0.193\,94$

3. El intervalo de control teórico para el número de imperfecciones por pieza es $\left(\lambda - 3\sqrt{\lambda}, \lambda + 3\sqrt{\lambda}\right) = \left(12 - 3\sqrt{12}, 1 + 3\sqrt{12}\right) = (1.\,607\,7, 22.\,392)$

   Como todos los elementos de la muestra están dentro de los límites de control, el proceso está bajo control manteniéndose el valor 12 como promedio del número de defectos de las piezas de tela.fabricadas.

   La gráfica siguiente muestra la carta de control, mostrando los intervalos de control y toda la muestra dentro de este intervalo:


c Chart for Col_1
UCL = 22.39
CTR = 12.00
LCL = 1.61
c
Observation
0
5
10
15
20
25
0
4
8
12
16
20
24

T. 7

Control de Variables

Ejercicio 89 *En un proceso industrial se controla la resistencia a la tensión de ciertas piezas metálicas. Para ello se ha medido la resistencia* (x_i) *de 30 muestras de 6 elementos cada una.obteniendose que la suma de las medias de las 30 muestras es 6000 y la suma de sus cuasidesviaciones 150.*

1. *Calcular, a partir de estas muestras los límites de control para la media y para la cuasidesviación.*
2. *Se ha concluido que el proceso está bajo control. Determinar el índice de capacidad si los límites de tolerancia son* 200 ± 5.
3. *¿Cuántas piezas defectuosas produce este proceso? (Se entiende que una pieza es defectuosa si sobrepasa los límites de tolerancia).*
4. *En un momento dado se desajusta el proceso y fabrica piezas con media 199, conservandose no obstante la varianza. ¿Cúal es la probabilidad de detectar el desajuste en la siguiente muestra de 6 elementos que se tome?.*

1. $\overline{\overline{X}} = \frac{6000}{30} = 200$; $\overline{s} = \frac{150}{30} = 5.0$; $\widehat{\sigma} = \frac{1}{c_4} \times \overline{s} = 1.0510 \times 5 = 5.255$
   Los límites de control para la media son: $\left(\overline{\overline{X}} - 3\frac{\widehat{\sigma}}{\sqrt{6}}, \overline{\overline{X}} + 3\frac{\widehat{\sigma}}{\sqrt{6}}\right) = \left(200 - 3\frac{5.255}{\sqrt{6}}, 200 + 3\frac{5.255}{\sqrt{6}}\right) = (193.564, 206.436)$
   Los límites de control para la cuasidesviación son
   $(\overline{s}B_3, \overline{s}B_4) = (5 \times 0.030, 5 \times 1.970) = (.15, 9.85)$
2. *Índice de capacidad* $= \frac{LT2 - LT1}{6\widehat{\sigma}} = \frac{205-195}{6\times 5.255} = 0.317\,158$
3. Calculamos la probabilidad de que las piezas tengan medidas fuera de los límites de tolerancia, usando una normal cuyos parámetros son los valores estimados de la muestra $N(200,\ 5.255)$

$1-P\,(195 < x < 205) = 1-P\left(\frac{195-200}{5.255} < z < \frac{205-200}{5.255}\right) =$
$= 1-P\,(-.951\,475 < z < .951\,475) = 1-(F(0.951475)-F(-0.951475))$
$= 0.341\,363$

4. Probabilidad de que la media de una muestra de 6 elementos de una distribución $N(199,\ 5.255)$ caiga fuera del intervalo $(193.\,564, 206.\,436)$.
   La media de esta muestra se distribuirá con una $N\left(199,\ \frac{5.255}{\sqrt{6}}\right)$
   $1-P\,(193.564 < x < 206.436) = 1-P\left(\frac{193.564-200}{\frac{5.255}{\sqrt{6}}} < z < \frac{206.436-200}{\frac{5.255}{\sqrt{6}}}\right) =$
   $1-P\,(-2.\,999\,99 < z < 2.\,999\,98)$
   $= 1-P\,(-3 < z < 3.) == 1-(F(3)-F(-3)) = 2.\,699\,8\times 10^{-3}$

Ejercicio 90 *Si la media del peso de unas latas de conservas es* $41.5\,gr$ *y la desviación típica es* $0.5\ gr$. *Se pide:*

1. *Hallar los límites de control teóricos para las medias muestrales si el número de elementos de cada muestra es $n = 5$*
2. *Hallar los límites de control para la cuasidesviación.*
3. *La siguiente tabla nos da los valores obtenidos para la media y la desviación típica de 20 muestras de tamaño $n = 5$ del mismo proceso*

$\overline{x}$	*41.9*	*41.3*	*42.1*	*41.6*	*41.8*	*42.3*	*41.4*	*41.6*	*41.8*	*42.0*
s	*0.8*	*0.2*	*0.3*	*0.7*	*0.9*	*0.1*	*0.4*	*0.5*	*0.6*	*0.2*

$\overline{x}$	*42.0*	*41.8*	*41.3*	*42.0*	*42.0*	*41.7*	*41.5*	*41.49*	*41.6*	*41.4*
s	*0.3*	*0.3*	*0.2*	*0.5*	*0.6*	*0.2*	*0.4*	*0.4*	*0.3*	*0.6*

¿Estos valores indican que el proceso está bajo control en media? ¿ Y en varianza?

1. $\left(\mu-3\frac{\sigma}{\sqrt{n}}, \mu+3\frac{\sigma}{\sqrt{n}}\right) = \left(41.5-3\frac{0.5}{\sqrt{5}}, 41.5+3\frac{0.5}{\sqrt{5}}\right) = (40.\,829\,2, 42.\,170\,8)$
2. $(\sigma B_5, \sigma B_6) = (0.5\times 0,\ \ 0.5\times 1.964) = (0,\ \ 0.982)$
3. La muestra 6^a está fuera de control en media. Ninguna muestra esta fuera de control en varianza.

Ejercicio 91 *Los diámetros de las arandelas que salen de una línea de fabricación siguen una distribución normal de media 0.5 cm y una desviación típica de 0.1 cm. Se pide:*

1. *Hallar los límites de control. teóricos 3-sigma para la media de muestras con 10 elementos.*

2. *Intervalo de control teórico para la varianza basado en la distribución chi-cuadrado*

3. *Se ha observado en un instante una muestra de 10 de estas arandelas. Las dimensiones de sus diámetros han sido: 0.4; 0.43; 0.6; 0.42; 0.7; 0.51; 0.61; 0.44; 0,62; 0.49. ¿Confirman estos valores el estado de control del proceso?*

4. *Calcular los límites de control para s basados en muestras de 10 elementos.*

5. *Si se analizan 100 muestras de 10 elementos cada una, ¿cuál es la probabilidad de que haya alguna fuera de control en media?*

1. $(0.5 - 3\frac{0.1}{\sqrt{10}}, 0.5 + 3\frac{0.1}{\sqrt{10}}) = (0.405\,13,\ 0.594\,87)$

2. $\chi^2_{9,0.00127} < \frac{nS^2}{\sigma^2} < \chi^2_{9,0.99873}$,de donde se deduce el intervalo para la varianza
   $\frac{\sigma^2 \chi^2_{9,0.00127}}{n} < S^2 < \frac{\sigma^2 \chi^2_{9,0.99873}}{n}, \quad \frac{0.1^2 \chi^2_{9,0.00127}}{10} < S^2 < \frac{0.1^2 \chi^2_{9,0.99873}}{10}$
   $\chi^2_{9,0.00127} = ^{\left(\chi^2_9\right)^{-1}} (0.00127) = 1.\,222\,5$
   $\chi^2_{9,0.99873} = ^{\left(\chi^2_9\right)^{-1}} (0.99873) = 27.\,253$
   Por lo tanto, el intervalo de confianza será:$\frac{0.1^2 1.2225}{10} < S^2 < \frac{0.1^2 27.253}{10}$
   $1.\,222\,5 \times 10^{-3} < S^2 < 2.\,725\,3 \times 10^{-2}$

   La media de la muestra es 0.519 y la varianza 0.0099. Tanto la media como la varianza están dentro de los límites de control. Así que se confirma el estado de control.

3. Los límites de control teóricos para la cuasidesviación s son:
   $(\sigma B_5,\ \sigma B_6) = (0.1 \times 0.276,\ 0.1 \times 1.669) = (0.027\,6,\ 0.166\,9)$

4. $1 - 0.997^{100} \quad = 0.\,259\,52$

Ejercicio 92 *Se ha observado en un instante una muestra de 10 chapas metálicas. Sus pesos en gramos han sido:* $40, 43, 60, 42, 70, 51, 61, 44, 62, 49$ *Hallar un intervalo de tolerancia que contenga el 90% de las piezas fabricadas con (95% de confianza).*

La media de la muestra es 52.2 y la cuasidesviacion 10.37. El intervalo de tolerancia es

$(\overline{x}-ks,\ \overline{x}+ks)=(52.2-2.839\times 10.37,\ \ 52.2+2.839\times 10.37)=$

$=(22.76,\ \ 81.64)$

Ejercicio 93 *Los diametros de las varillas fabricadas en una maquina es una característica importante en su calidad. La siguiente tabla muestra los valores de y de R para 20 muestras de 5 varillas cada una. Las especificaciones de las varillas son 0.5035±0.0010 pulgadas. Los valores dados en la tabla son las últimas tres cifras de la medida. Es decir 34.2 significa 0.50342.*

$Muestra$	$\overline{x}$	R	$Muestra$	$\overline{x}$	R
1	0.50342 34.2	0.0003 3	*11*	35.4 0.50354	0.0008 8
2	31.6	4	*12*	34.0	6
3	31.8	4	*13*	36.0	4
4	33.4	5	*14*	37.2	7
5	35.0	4	*15*	35.2	3
6	32.1	2	*16*	33.4	10
7	32.6	7	*17*	35.0	4
8	33.8	9	*18*	34.4	7
9	34.8	10	*19*	33.9	8
10	38.6	4	*20*	34.0	4

1. *Hallar la media de las muestras, el rango medio y revisar si fuera necesario los límites de control.*
2. *Calcular índice de capacidad del proceso.*
3. *¿Qué porcentajes de defectos está produciendo este proceso*

1. La media de la muestra es 0.50343 y la media de los rangos 0.000565. los límites de control provisionales son:

   $(0.50343-3\frac{0.000565}{2.326\sqrt{5}},0.503409+3\frac{0.000565}{2.326\sqrt{5}})=(0.5031,\ \ 0.50373)$

   La muestra n^o 10 queda fuera de estos límites. Eliminándola, la media muestral es ahora 0.503409 y el rango medio 0.000573684. Hallando de nuevo los límites de control se obtiene:

   $(0.503409-3\frac{0.000573684}{2.326\sqrt{5}},0.503409+3\frac{0.000573684}{2.326\sqrt{5}})=(0.50308,\ \ 0.50374)$

   Estos límites de control ya son válidos.

   La estimación de la desviación típica es:$\frac{0.000573684}{2.326}=2.4664\times 10^{-4}$

2. El índice de capacidad es por tanto:
   $IC = \frac{LST-LIT}{6\sigma} = \frac{0.5045-0.5025}{6\times 2.\,466\,4\times 10^{-4}} = 1.\,351\,5$
3. $P(0.5025 < x < 0.5045) = 0.999\,99 - 7.\,879\,7\times 10^{-5} = 0.999\,91.$

   La probabilidad de producir varillas fuera de los límites de especificación es $1 - 0.999\,91 = 0.000\,09$. La producción de piezas defectuosas es prácticamente nula

Ejercicio 94 *Supongamos que se usa un diagrama de control 3-sigma para un proceso distribuido normalmente. Cada 2 horas se toman muestras de 30 elementos y se marca el punto correspondiente en el diagrama. Hallar el número esperado de muestras que se habrán inspeccionado hasta que un punto esté fuera de los límites de control.*

La probabilidad de que una muestra tenga la media fuera de los límites de control es 0.0027

> Por tanto por promedio se deben de inspeccionar $1 \times 0.0027 + 2 \times 0.9973 \times 0.0027 + 3 \times 0.9973^2 \times 0.0027 + ... = \sum_{k=1}^{\infty} k \times 0.9973^{k-1} 0.0027 = 370.\,37$ muestras.

Ejercicio 95 *Un proceso de fabricación varillas de aluminio está bajo control. Las especificaciones indican que los diametros de la varillas siguen una distribución normal N(1.25, 0.01). Para realizar el control de calidad se toman muestras de 5 elementos cada hora.*

1. *¿Cuáles serán los límites de control 3-sigma.?*
2. *¿Cuál es la probabilidad de obtener una muestra fuera de control en una prueba?*
3. *¿Cuál es la probabilidad de obtener 3 muestras fuera de control en 10 pruebas?*
4. *¿Cuál es la probabilidad de obtener 1 muestras fuera de control en 100 pruebas?*

1. $(1.25 - 3\frac{0.01}{\sqrt{5}}, 1.25 + 3\frac{0.01}{\sqrt{5}}) = (\ 1.\,236\,6, \quad 1.\,263\,4\ )$
2. 0.0027
3. $\binom{10}{3} 0.0027^3 0.9973^7 = 2.\,317\,7 \times 10^{-6} = 0.0000023$
4. $\binom{100}{1} 0.0027^1 0.9973^{99} = 0.\,206\,6$

Ejercicio 96 *Si la media de la medidas del diametro de unas varillas es* $\mu = 4.2$ *y la desviación típica es* $\sigma = 0.05$

Se pide:

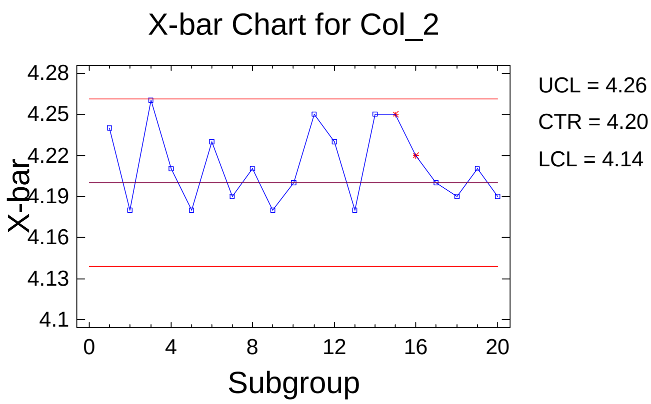
1. *Hallar los límites de control teóricos si el número de elementos de cada muestra es n=6*
2. *Hallar los límites de control 3-sigma para la desviación típica*
3. *La siguiente tabla nos da los valores obtenidos para la media y la desviación típica de 20 muestras de tamaño n=6 del mismo proceso*

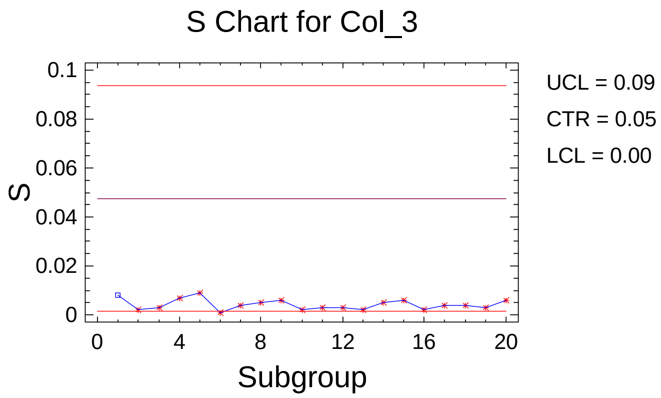
$\overline{x}$	*4.24*	*4.18*	*4.26*	*4.21*	*4.18*	*4.23*	*4.19*	*4.21*	*4.18*	*4.20*
s	*0.008*	*0.002*	*0.003*	*0.007*	*0.009*	*0.001*	*0.004*	*0.005*	*0.006*	*0.002*

4.25	*4.23*	*4.18*	*4.25*	*4.25*	*4.22*	*4.20*	*4.19*	*4.21*	*4.19*
0.003	*0.003*	*0.002*	*0.005*	*0.006*	*0.002*	*0.004*	*0.004*	*0.003*	*0.006*

¿Estos valores indican que el proceso está bajo control en media? ¿ Y en varianza?. Hacer ambas gráficas de control

1. $(4.2 - 3\frac{0.05}{\sqrt{6}}, 4.2 + 3\frac{0.05}{\sqrt{6}}) = \left(\ 4.138\,8,\quad 4.261\,2\ \right)$
2. $(\sigma(c_2 - 3c_3), \sigma(c_2 - 3c_3)) = (0.05(0.9515 - 3 \times 0.3075), 0.05(0.9515 + 3 \times 0.3075))$
   $= (0.05 \times 0.029\,, 0.05 \times 1.874) = (0.001\,45, 0.093\,7)$
3. Esta bajo control en media, pero no varianza. La muestra n^o 6 tiene la desviación típica 0.001, que es menor que el límite inferior de control (0.001 45).
4. Las gráficas de control para la media y la desviación típica se muestran a continuación:





Se observa en la gráfica de control de la desviación típica que el proceso tiene menos variabilidad de la prevista. Como esta circunstancia suele ser favorable, sería conveniente evaluar un nuevo valor,más pequeño, para la especificación de la desviación típica y realizar una observación del proceso para ver si se puede localizar el motivo de esta mejoría

Ejercicio 97 *Se supone que un proceso produce piezas cuyas medidas sigue una distribución normal con* $\mu = 100,\ \sigma = 2$.

1. *Calcular los límites de control de calidad teóricos, para muestras de tamaño 5.*

2. *Se ha realizado un control de calidad para verificar la válidez de estos parámetros, obteniendose los resultados siguientes con muestras de tamaño 5*

Muestra	*media*	*recorrido*	*Muestra*	*media*	*recorrido*
1	*99.7*	*3.1*	*11*	*98.4*	*3.1*
2	*99.8*	*3.4*	*12*	*98.5*	*2.8*
3	*100.0*	*3.3*	*13*	*97.9*	*2.9*
4	*99.8*	*3.6*	*14*	*98.5*	*3.2*
5	*99.9*	*3.0*	*15*	*100.8*	*3.1*
6	*99.7*	*3.2*	*16*	*100.5*	*3.3*
7	*100.1*	*3.1*	*17*	*99.4*	*3.4*
8	*100.2*	*2.9*	*18*	*99.9*	*3.0*
9	*99.3*	*1.9*	*19*	*97.5*	*3.5*
10	*99.7*	*3*	*20*	*99.2*	*3.3*

¿Se deben cambiar los límites de control?

3. *Si se consideran defectuosas las piezas que estan fuera del intervalo (94,106) ¿Qué proporción de piezas defectuosas produce el proceso?*

1. Los límites de control teórico son $(100-3\frac{2}{\sqrt{5}}, 100+3\frac{2}{\sqrt{5}}) = (97.317, 102.68)$.

2. No es necesario cambiar los límites de control para la media, porque todos las medias quedan dentro de este intervalo. Para ver si hay que variar la varianza estimo la varianza a partir de la media de los rangos: $\widehat{\sigma} = d\overline{R} = \frac{1}{2.3226}3.105 = 1.3369$

   Los límites de confianza al 95%a partir de la muestra para σ^2 son $\frac{(n-1)s^2}{\left(\chi_4^2\right)^{-1}(0.975)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\left(\chi_4^2\right)^{-1}(0.025)}$; $\frac{4\times 1.3369^2}{11.143} \leq \sigma^2 \leq \frac{4\times 1.3369^2}{0.48442}$; $0.64159 \leq \sigma^2 \leq 14.758$.

   Como la varianza teórica es 4, entra dentro de este intervalo de confianza, por tanto la muestra es congruente con los datos teóricos. Además se puede comprobar que está bajo control.

3. Como no se descarta los datos teóricos, las medidas de las piezas siguen una distribución N(100,2). la probabilidad de piezas aceptables será:

   $P(94 < x < 106) = 0.99865 - 1.3499 \times 10^{-3} = 0.9973$

   por tanto la proporción de defectuosos es $1 - 0.9973 = 0.0027$

Ejercicio 98 *La longitud del encendedor de cigarrillos de un automóvil es controlada mediante el empleo de gráficos de control para la media y para el recorrido. La siguiente tabla proporciona las medidas de la longitud para 20 muestras de tamaño 4.*

	Observaciones					*Observaciones*			
nº	*1*	*2*	*3*	*4*	*nº*	*1*	*2*	*3*	*4*
1	*5.15*	*5.10*	*5.08*	*5.09*	*11*	*5.13*	*5.08*	*5.09*	*5.05*
2	*5.14*	*5.14*	*5.10*	*5.06*	*12*	*5.10*	*5.15*	*5.08*	*5.10*
3	*5.09*	*5.10*	*5.09*	*5.11*	*13*	*5.08*	*5.12*	*5.14*	*5.09*
4	*5.08*	*5.06*	*5.09*	*5.13*	*14*	*5 .15*	*5.12*	*5.14*	*5.05*
5	*5.14*	*5.08*	*5.09*	*5.12*	*15*	*5.13*	*5.16*	*5.09*	*5.05*
6	*5.09*	*5.10*	*5.07*	*5.13*	*16*	*5.14*	*5.08*	*5.08*	*5.12*
7	*5.15*	*5.10*	*5.12*	*5.12*	*17*	*5.08*	*5.10*	*5.16*	*5.09*
8	*5.14*	*5.16*	*5.11*	*5.10*	*18*	*5.08*	*5.14*	*5.10*	*5.09*
9	*5.11*	*5.07*	*5.16*	*5.10*	*19*	*5.13*	*5.15*	*5.10*	*5.08*
10	*5.11*	*5.14*	*5.11*	*5.12*	*20*	*5.09*	*5.07*	*5.15*	*5.08*

1. *Hallar los límites superiores e inferiores de control, para la media y el rango de cada muestra y eliminando, si es necesario, las muestras fueras de control.*

2. *Si el intervalo de tolerancia es (5.05, 5.15) calcula el índice de capacidad y estima la proporción de encendedores que quedarían fuera de este intervalo de tolerancia.*

1. Hay que calcular la media y los rangos de cada una de las muestras

n^o	Media	Rango	n^o	Media	Rango
1	5.105	0.07	11	5.0875	0.08
2	5.11	0.08	12	5.1075	0.07
3	5.0975	0.02	13	5.1075	0.06
4	5.09	0.07	14	5.115	0.1
5	5.1075	0.06	15	5.1075	0.11
6	5.0975	0.06	16	5.105	0.06
7	5.1225	0.05	17	5.1075	0.08
8	5.1275	0.06	18	5.1025	0.06
9	5.11	0.09	19	5.115	0.07
10	5.12	0.03	20	5.0975	0.08

La media de la muestra es 5.107 y la media de los rangos 0.068. los límites de control provisionales son:

$(5.107 - 3\frac{0.068}{2.059\sqrt{4}}, 5.107 + 3\frac{0.068}{2.059\sqrt{4}}) = (5.057\,5,\ 5.156\,5\ ).$

No hay muestras fuera de control en media.

2. Los límites de control para los rangos son $(D_3\overline{R}, D_4\overline{R}) = (0\times 0.068, 2.282\times 0.068) = (0, 0.155\,18)$.

No hay muestras fuera de control

La estimación de la desviación típica es:$\frac{0.068}{2.059} = 3.302\,6 \times 10^{-2}$

3. El índice de capacidad es por tanto:

$IC = \frac{LST-LIT}{6\sigma} = \frac{5.15-5.05}{6\times 3.302\,6\times 10^{-2}} = 0.504\,65$

4. $P(5.05 < x < 5.15) = 0.925\,50 - 2.788\,2 \times 10^{-2} = 0.897\,62$.

5. Por tanto quedarían fuera de este intervalo de confianza una proporción de encendedores $1 - 0.897\,62 = 0.102\,38$

Ejercicio 99 *En una empresa envasadora de cervezas se han realizado gráficos de control, basados en muestras de 4 botellas, para la media y el rango del contenido en cerveza de las botellas que comercializa. Estos diagramas han dado lugar a los siguientes datos: a)Para el diagrama de la media LSC=330.99, linea central 328, LIC=325.01 b) Para el diagrama del rango LSC = 9.4 Linea central 4.1, LIC=0. Se ha confirmado que el proceso está bajo control.*

1. *Calcular la desviación típica estimada por el proceso.*
2. *Si usamos estos los datos del diagrama de rango para realizar un diagrama de control para la desviación típica, ¿Que aspecto presentaría?*
3. *Si se supone que estas botellas quieren venderse como botellas de un tercio y que las especificaciones que desea dar el fabricante para el contenido de las mismas son* 333 ± 8 cm^3 *¿Que porcentaje de botellas resultarían fuera de estas especificaciones?¿Que correción se debería hacer en el proceso de envasado para disminuir este porcentaje?*

1. $\widehat{\sigma} = \frac{1}{d_2}\overline{r} = \frac{1}{2.059} \times 4.1 = 1.\,991\,3$
2. Empleando el valor de la desviación típica estimada como valor para $\frac{1}{c_2}\overline{S}$ se obtiene un valor aproximado para $\overline{S}$, $\overline{S} = c_2\widehat{\sigma} = 0.7979 \times 1.\,991\,3 = 1.\,588\,9$, y para $\overline{s} = \overline{S}\sqrt{\frac{n}{n-1}} = 1.5889 \times \sqrt{\frac{4}{3}} = 1.\,834\,7$. El intervalo de control para la desviación típica muestral es:

   $\left(\overline{S}B_3,\ \overline{S}B_4\right) = (1.5889 \times 0, 1.5889 \times 2.266) = (0, 3.\,600\,4)$

   El intervalo de control para la cuasidesviación s de la muestra es:

   $(\bar{s}B_3,\ \bar{s}B_4) = (1.\,83479 \times 0,\ 1.\,8347 \times 2.266) = (0,\ 4.\,157\,4)$
3. Los parámetros para la media y la desviación típica del proceso de fabricación son 328 y 1.9913 respectivamente.

   La proporción de botellas que cumplen las especificaciones sería:

   $P(325 < X < 341) = F(341) - F(325) = 1 - 0.993\,979\,2 = 1.0 - 0.5 = 0.5$

   Sólo la mitad de las botellas cumplen las especificaciones, así que el 50% de ellas no las cumplen.

   Para arreglar este alto porcentaje de botellas fuera de las especificaciones, se debería intentar ajustar la maquinaria para que la media del proceso, que ahora es 325, se acerque lo más posible al valor central, 333 cm^3, de las especificaciones. Si se consiguiera ajustarlo perfectamente, suponiendo que se conserve la desviación típica, la proporción de botellas disconformes sería::

   $1 - P(325 < X < 341) = 1 - (1.\,000\,0 - 6.\,596\,3 \times 10^{-2}) = 6.\,596\,3 \times 10^{-2}$

T. 8

Control de Recepción

Ejercicio 100 *Un producto se produce en lotes de 25 piezas. El procedimiento de inspeción consiste en seleccionar una muestra de 5 elementos de cada lote. El lote se acepta sólo si no aparece ningun elemento defectuoso en la muestra.*

1. *¿Cual es la probabilidad de aceptar un lote que contenga 3 elementos defectuosos?*
2. *¿Sería adecuado usar la aproximación bimomial para calcular esta probabilidad? ¿Y si los lotes fueran de 200 piezas.*
3. *Supongamos que un lote de 25 piezas contiene 5 defectuosas. El procedimiento de muestreo va a consistir en seleccionar una muestra de algunas muestras de este lote. Si la muestra contiene algún elemento defectuoso se rechazará el lote. Si se quiere que la probabilidad de rechazar este lote sea al menos 0.95 ¿Cuántos elementos ha de tener la muestra?*

1. Como los lotes tienen pocos elementos, usamos la distribución hipergeométrica (nd significa el número de defectos de la muestra):
2. $P(nd = 0) = \frac{\binom{3}{0}\binom{22}{5}}{\binom{25}{5}} = 0.495\,65$

   Usando la distribución binomial se obtiene el valor:

   $P(nd = 0) = \binom{5}{0}\left(\frac{3}{25}\right)^0\left(\frac{22}{25}\right)^5 = 0.527\,73$

   Se distingue del anterior $0.527\,73 - 0.495\,65 = 0.032\,08$, lo que supone un error relativo $\frac{0.032\,08}{0.495\,65} = 6.\,472\,3 \times 10^{-2}$, equivalente prácticamente al 6.5%.

   Si los lotes fueran grandes, de 200 elementos los valores de la probabilidad serían

$\frac{\binom{3}{0}\binom{197}{5}}{\binom{200}{5}} = 0.926\,50$; $\binom{5}{0}\left(\frac{3}{200}\right)^0\left(\frac{197}{200}\right)^5 = 0.927\,22$

En este caso el error relativo del orden del 0.078%. En este caso es poco importante la diferencia entre los valores obtenidos por cada una de las distribuciones para la probabilidad de aceptar el lote.

3. La probabilidad de rechazar el lote usando una muestra de n elementos será: $1 - \frac{\binom{5}{0}\binom{20}{n}}{\binom{25}{n}}$.

   Dando valores sucesivamente a n se obtiene

   $1 - \frac{\binom{5}{0}\binom{20}{5}}{\binom{25}{5}} = 0.708\,19$; $1 - \frac{\binom{5}{0}\binom{20}{6}}{\binom{25}{6}} = 0.781\,14$

   $1 - \frac{\binom{5}{0}\binom{20}{7}}{\binom{25}{7}} = 0.838\,74$; $1 - \frac{\binom{5}{0}\binom{20}{8}}{\binom{25}{8}} = 0.883\,53$

   $1 - \frac{\binom{5}{0}\binom{20}{9}}{\binom{25}{9}} = 0.917\,79$; $1 - \frac{\binom{5}{0}\binom{20}{10}}{\binom{25}{10}} = 0.943\,48$

   $1 - \frac{\binom{5}{0}\binom{20}{11}}{\binom{25}{11}} = 0.962\,32$

   Por tanto hay que inspeccionar al menos 11 piezas.

Ejercicio 101 *Entre un vendedor y un cliente se ha acordado un plan de muestreo que consiste en que el comprador aceptará los lotes si en una muestra de 100 piezas el número de defectuosas no es mayor de 10. En caso contrario los rechazará.*

1. *¿Cuál es la probabilidad de aceptar un lote con un número de defectuosas del 15%? ¿ Y del 25%?*

2. *Da la expresión analítica para la curva característica de este plan de muestreo.*

1. Si el promedio de defectuosas fuera 15% la distribución del n^o de defectuosas cada 100 sería una B(100, 0.15). Aproximamos con la normal de *media* $= np = 15$ y *desviación típica* $\sqrt{npq} = \sqrt{100 \times 0.15 \times 0.85} = 3.\,570\,7$

   En este caso

   $P(n^o def. \leq 10) = P(x < 10.5) = P(z < \frac{10.5-15}{3.57}) = P(z < -1.\,260\,5) = 0.103\,83.$

   Si el número de defectos fuera del 25% la media sería 25 y la desviación típica valdría 4.33 por lo tanto

   $P(n^o def. \leq 10) = P(x < 10.5) = P(z < \frac{10.5-25}{4.33}) = P(z < -3.35) \approx 0$

2. La curva característica de un plan de muestreo representa la probabilidad de aceptar el lote frente a la proporción real de defectuosos en él.

   $P_A(p)$ = Probabilidad de aceptar un lote con una proporción p de defectuosos $= \sum_{i=1}^{10} \binom{100}{i} p^i (1-p)^{100-i}$

Ejercicio 102 *Un plan de muestreo doble consiste en dos fases. En la primera se inspeccionan 10 elementos de una muestra. Si no hay defectuosos se acepta el lote, Si el número de defectuosos es 3 o más de 3 se rechaza. En otro caso se toma una segunda muestra de 20 elementos, Si entre los 30 elementos inspeccionados en total el número de defectuosos no supera los 3, el lote se acepta, en otro caso se rechaza.¿Cuál es la probabilidad de aceptar un lote con una proporción de piezas defectuosas de 10%? ¿Y con una proporción de elementos defectuosos del 20%?*

La probabilidad de aceptar el lote es la prob de que no haya defectuosos en la 1^a fase más la probabilidad que haya 1 o 2 defectuosos en la 1^a fase y el número total de defectuosos no supere los 3:

Si la proporción de piezas defectuosas es del 10% la probabilidad de aceptar el lote es:

$\binom{10}{0}0.1^0 0.9^{10} + \binom{10}{1}0.1^1 0.9^9 \left[\binom{20}{0}0.1^0.9^{20} + \binom{20}{1}0.1^1.9^{19} + \binom{20}{2}0.1^2.9^{18}\right] +$

$+\binom{10}{2}0.1^2 0.9^8 \left[\binom{20}{0}0.1^0 0.9^{20} + \binom{20}{1}0.1^1.9^{19}\right] = 0.348\,68 + 0.387\,42 \times 0.676\,93 + 0.193\,71 \times 0.391\,75 = 0.686\,82$

Si la proporción real de elementos defectuosos fuera del 20%, la probabilidad de aceptar el lote es:

$\binom{10}{0}0.2^0 0.8^{10} + \binom{10}{1}0.2^1 0.8^9 \left[\binom{20}{0}0.2^0.8^{20} + \binom{20}{1}0.2^1.8^{19} + \binom{20}{2}0.2^2.2^{18}\right] +$
$+\binom{10}{2}0.2^2 0.8^8 \left[\binom{20}{0}0.2^0 0.8^{20} + \binom{20}{1}0.2^1.8^{19}\right] =$
$= 0.107\,37 + 0.268\,44 \times \; 6.917\,5 \times 10^{-2} + 0.301\,99 \times 6.917\,5 \times 10^{-2} = 0.146\,83$

Ejercicio 103 *Se ha establecido un plan de muestreo en recepción para lotes de una gran cantidad de arandelas. Las características de este plan son las siguientes: Se inspeccionará una muestra de 100 de estas arandelas. Se aceptará el lote si el número de piezas defectuosas no es superior a 7. En caso contrario se rechazará el lote.*

1. *Si porcentaje real de defectos de un lote es del 5%, ¿ Cuál es la probabilidad de aceptar el lote?. ¿Cuál es la probabilidad de rechazar el lote si el porcentaje de defectos fuera del 10%?*

2. *Calcular las ordenadas de la curva característica del plan para los porcentajes de defectos 5%, 20% y 25% .*

3. *Como se sabe un plan de muestreo se establece usando 4 parámetros (P_A, α, P_R, β), donde α es el riesgo del vendedor y β es el riesgo del comprador. Si P_A es 0.06 y β es 0.05, calcular los valores de α, P_R que corresponden a este plan de muestreo*

4. *Da la expresión y la representación gráfica de la curva caractéristica de este plan de muestreo.*

1. La probabilidad de aceptar el lote si el porcentajes de defectos fuera del 5% sería

   $P(nd. \leq 7)$ se rige por una $B(100, 0.05)$. calculando esta probabilidad por medio de la aproximación a la normal se tiene que $P(x < 7.5)$ para una $N(100 \times 0.05, \sqrt{100 \times 0.05 \times 0.95}))$ es $P(z < \frac{7.5-100\times0.05}{\sqrt{100\times0.05\times0.95}}) =$

   $P(z < 1.1471) = 0.87433$.

   Usando la distribución binomial directamente habríamos obtenido:

   $\sum_{i=0}^{7} \binom{100}{i} 0.05^i 0.95^{100-i} = 0.87204$

   La probabilidad de rechazar el lote si el porcentaje de defectos fuera del 10% se puede calcular del modo siguiente:

   $P(n^o\ def. > 7)$ en una $B(100, 0.1)) =$

   $P(x > 7.5\,)$ para una $N(10, 3)) = P(z > \frac{7.5-10}{3}) = P(z > \frac{7.5-10}{3} = -.83333) = 0.79767$

2. La ordenada de la curva característica del plan es la probabilidad de aceptar el lote para un porcentaje dado de defectos en él. Para el porcentaje de defectos del 5% es, tal como ya hemos calculado en el primer apartado es 0.87.

   Para una proporción de defectos del 20%

   $P(z < \frac{7.5-100\times0.2}{\sqrt{100\times0.2\times0.8}}) = P(z < -3.125) = 0.8.8903 \times 10^{-4}$

   Si usamos directamente la distribución binomial resulta:

   $\sum_{i=0}^{7} \binom{100}{i} 0.2^i 0.8^{100-i} = 2.7699 \times 10^{-4}$

   Para una proporción de defectos en el lote del 25% es:

   $P(z < \frac{7.5-100\times0.25}{\sqrt{100\times0.25\times0.75}}) = P(z < -4.0415) = 2.6555 \times 10^{-5}$

   que es prácticamente nula.

3. $\alpha =$ Riesgo del vendedor $= P(\text{rechazar un lote con } P = P_A)$

   Sustituyendo los valores del problema se obtiene:

   $\alpha =$ Riesgo del vendedor $= P(\text{rechazar un lote con } P = 0.06)$

   $P(nd. > 7 \text{ en una binomial } B(100, 0.06)) =$

   $P(x > 7.5 \text{ para una } N(100 \times 0.06, \sqrt{100 \times 0.06 \times 0.94})) =$

   $P\left(z > \frac{7.5-6}{2.3749}\right) = P\left(z > .631\,61\right) = 1 - P\left(z \leq .631\,61\right) = 0.263\,82.$ Hemos obtenido el valor de α.

   Hallamos el valor de P_R a partir de la expresión:

   $\beta =$ Riesgo del comprador $= P(\text{aceptar un lote con } P = P_R)$

   $0.05 = P(\textit{aceptar un lote con } P = P_R) =$

   $P(nd. \leq 7) \text{ en una binomial } B(100P_R, \sqrt{100P_R\,(1-P_R)})) =$

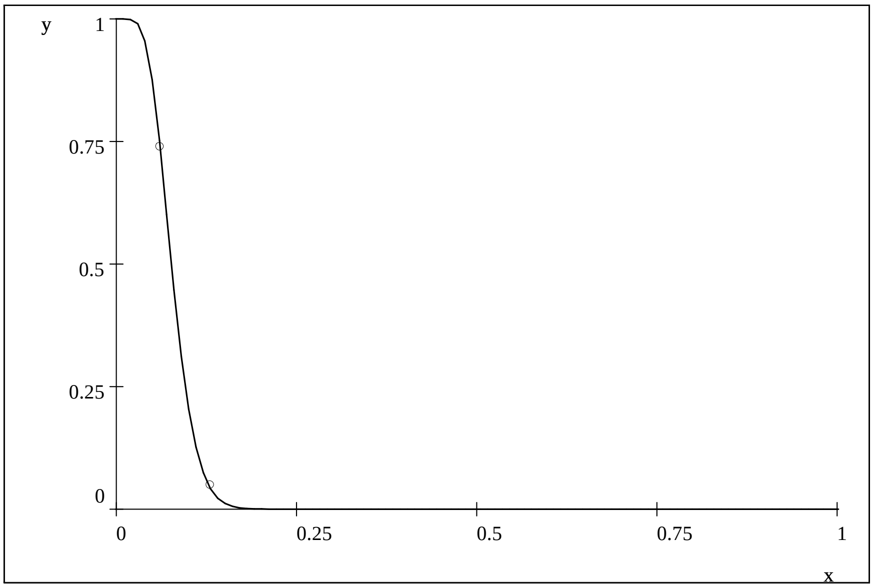
   $= P\left(z < \frac{7.5-100P_R}{\sqrt{100P_R(1-P_R)}}\right) = 0.05 \Rightarrow \frac{7.5-100P_R}{\sqrt{100P_R(1-P_R)}} = -1.644\,9$

   Resolviendo esta última ecuación $\frac{7.5-100P_R}{\sqrt{100P_R(1-P_R)}} = -1.644\,9$ se obtiene $P_R = 0.130\,39$

4. La expresión de la curva característica de este plan de muestreo es:

   $P_A(p) = \sum_{i=0}^{7} \binom{100}{i} p^i \,(1-p)^{100-i}$

   La representación gráfica es la siguiente:



Están también representados los puntos que corresponden a:

$$(P_A,\ 1-\alpha) = (0.06,\ 0.74) \text{ y } (P_R,\ \beta) = (0.13, 0.05)$$

Ejercicio 104 *Diseñar un plan de muestreo secuencial truncado siendo* $p_A = 0.05$, $\alpha = 0.05$ *y* $p_R = 0.15$ $\beta = 0.10$ *y* $k = 5$.

Indicar qué decisión hay que tomar en las siguientes situaciones:

1. *Se han inspeccionado 30 piezas y se han encontrado 2 defectuosas*

2. *Se han inspeccionado 50 piezas y se han encontrado 3 defectuosas*

3. *Se han inspeccionado 61 piezas y se han encontrado 6 defectuosas*

Hay que utilizar las siguientes expresiones:

$$h_1 = \frac{\log\left(\frac{1-\alpha}{\beta}\right)}{\log\frac{P_R}{P_A} + \log\frac{1-P_A}{1-P_R}}, \quad h_2 = \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\log\frac{P_R}{P_A} + \log\frac{1-P_A}{1-P_R}}, \quad S = \frac{\log\frac{1-P_A}{1-P_R}}{\log\frac{P_R}{P_A} + \log\frac{1-P_A}{1-P_R}}$$

$$h_1 = \frac{\log\left(\frac{0.95}{0.10}\right)}{\log\frac{0.15}{0.05} + \log\frac{0.95}{0.85}} = 1.8608$$

$$h_2 = \frac{\log\left(\frac{0.90}{0.05}\right)}{\log\frac{0.15}{0.05} + \log\frac{0.95}{0.85}} = 2.3891$$

$$S = \frac{\log\frac{0.95}{0.85}}{\log\frac{0.15}{0.05} + \log\frac{0.95}{0.85}} = 9.1934 \times 10^{-2}$$

La recta de aceptación es $y = 9.1934 \times 10^{-2}x - 1.8608$
la recta de rechazo es $y = 9.1934 \times 10^{-2}x + 2.3891$
las rectas de acotación son
$x = 3 \times \frac{1.8608}{9.1934 \times 10^{-2}} = 60.722$, $y = kh_1 = 3 \times 1.8608 = 5.5824$.

Es conveniente calcular también los puntos de intersección de las rectas paralelas con las de truncamiento, que son $(34.735, 5.5824,)$, y $(60.722, 3.7216,)$

1. En el primer caso el número de elementos de la muestra es $30 < 34.735$, así que sus límites son las dos rectas paralelas. Hallamos sus ordenadas para $x = 30$:

   El valor de la ordenada de recta de aceptación es $y = 9.1934 \times 10^{-2} \times 30 - 1.8608 = 0.89722$

   El valor de la ordenada de la recta de rechazo es $y = 9.1934 \times 10^{-2} \times 30 + 2.3891 = 5.1471$

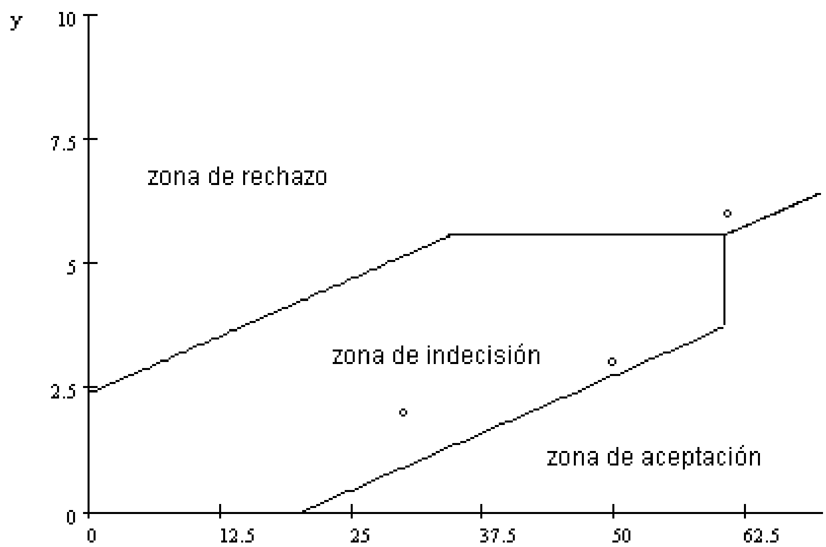
   Como el número de defectuosas es 2 que está entre ambos valores, no podemos decidir aún, y tenemos que continuar muestreando.

2. Como $50 > 34.735$ tenemos que considerar la recta inferior y la truncamiento que es $y = 5.5824$.

   Considerando la recta inferior obtenemos $y = 9.193\,4 \times 10^{-2} \times 50 - 1.8608 = 2.735\,9$. Como el número de defectuosas, 3 está entre estas dos cotas se debe seguir muestreando.

3. En este último caso comparamos hemos pasado el truncamiento $x = 60.722$, así que ya hay que tomar una decisión. para ello vamos a emplear la recta $y = 9.193\,4 \times 10^{-2}x = 9.193\,4 \times 10^{-2} \times 61 = 5.608\,0$. Como el número de defectuosos es 6 queda en la zona de rechazo y por tanto no aceptaríamos el lote.

   La siguiente gráfica representa la región de indecisión del plan. La parte superior es de rechazo, la inferior de aceptación. Tambien estan representados los puntos que corresponden a los tres casos propuestos.



Unidad Temática III

PROBLEMAS DE FIABILIDAD

T. 9

Fiabilidad y Fallos

Ejercicio 105 *Comprueba la propiedad de falta de memoría, ($P(\tau > t + h/\tau > t) = P(\tau > h)$, de la distribución exponencial y que esta propiedad no se cumple si la distribución es uniforme con función de densidad:*

$$\begin{cases} f(t) = 0.1 & \textit{si } 0 \le t \le 10 \\ f(t) = 0 & \textit{en el resto} \end{cases}$$

La función de fiabilidad de la exponencial es $R(t) = e^{-\lambda t}$

$P(\tau > t + h/\tau > t) = P(\tau > h)$

$P(\tau > t + h/\tau > t) = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h}$

$P(\tau > h) = e^{-\lambda h}$

Por tanto la igualdad se cumple para la distribución exponencial

Probamos ahora que no se cumple para la uniforme.

$P(\tau > t + h/\tau > t) = P(\tau > h)$

$P(\tau > t + h/\tau > t) = \frac{(10-t-h)0.1}{(10-t)0.1} = \frac{10-t-h}{10-t} = 1 - 0.1\,(t + h)$

$P(\tau > h) = (10 - h)0.1 = 1 - 0.1h$

Ejercicio 106 *Calcular las funciones de fiabilidad, infiabilidad y de densidad que corresponden a una función tasa de fallo $h(t) = t$ con $t \ge 0$.*

$R(t) = \exp\left(-\int_0^t h(t)\,dt\right) = \exp\left(-\int_0^t t\,dt\right) = e^{-t^2/2},\ t \ge 0$

$F(t) = 1 - R(t) = 1 - e^{-t^2/2},\ \ t \ge 0$

$f(t) = F'(t) = \frac{d}{dt}(1 - e^{-t^2/2}) = te^{-t^2/2},\ \ t \ge 0$

Ejercicio 107 *El tiempo de vida de unos dispositivos sigue una distribución Normal de media 5000 horas y desviación típica 500 horas.*

1. *Calcular la función de fiabilidad y la probabilidad de que uno de estos dispositivos dure al menos 4500 horas.*

2. *Si se sabe que uno de estos dispositivos ya ha durado 4500 horas, ¿Cuál es la probabilidad de que dure por lo menos 500 horas más?*

1. $R(t) = P(\tau > t) = 1 - P(\leq t) = 1 - \int\limits_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \int\limits_{t}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$

   $R(4500) = P(\tau > 4500) = P\left(\frac{\tau-5000}{500} > \frac{4500-5000}{500}\right) = P(z > -1) = P(z < 1) = 0.841\,34$

2. $P(\tau > 5000/\tau > 4500) = \frac{P(\tau>5000)}{P(\tau>4500)} = \frac{P(z>0)}{P(z>-1)} = \frac{0.5}{0.8413} = 0.594\,32$

Ejercicio 108 *El tiempo de vida de unos dispositivos sigue una distribución exponencial de media 5000 horas*

1. *Calcular la función de fiabilidad y la probabilidad de que uno de estos dispositivos dure al menos 4500 horas.*

2. *Si se sabe que uno de estos dispositivos ya ha durado 4500 horas, ¿Cuál es la probabilidad de que dure por lo menos 500 horas más?*

1. La tasa de fallo es $\frac{1}{5000} = 0.000\,2$, por tanto la función de densidad es: $f(t) = 2 \times 10^{-4} e^{-2\times10^{-4}t}$; $F(t) = \int\limits_{0}^{t} 2 \times 10^{-4} e^{-2\times10^{-4}t} dt = 2 \times 10^{-4} \frac{e^{-2\times10^{-4}t}}{-2\times10^{-4}}\Big|_0^t = -e^{-2\times10^{-4}t} - (-1) = 1 - e^{-2\times10^{-4}t}$

   $R(t) = 1 - F(t) = e^{-2\times10^{-4}t}$

   $R(4500) = e^{-2\times10^{-4}\times4500} = 0.406\,57$

2. $P(\tau > 5000/\tau > 4500) = \frac{P(\tau>5000)}{P(\tau>4500)} = \frac{e^{-2\times10^{-4}\times5000}}{0.40657} = 0.904\,84$

Ejercicio 109 *La función de densidad (en horas) de la duración en funcionamiento de ciertos componentes es* $f(t) = 9te^{-3t}$ *para* $t > 0$.

1. *Hallar la vida media de dicho componente*

2. *Hallar la probabilidad de que un componente dure al menos 1 hora.*

3. *Si al principio había 10000 elementos, ¿cuantos se esperan que sobrevivan despues de la primera hora*

4. *Hallar la probabilidad de que un componente que haya durado ya 1 hora, resista todavía sin fallar por lo menos una hora más.*

5. *Hallar la función tasa de fallo y su valor para $t = 1$ por hora y por minuto*

6. *Calcular la probabilidad de que sobrevivan un minuto los que ya han durado una hora.*

1. Comprobamos en primer lugar que la función dada es una función densidad, puesto que $f(t) > 0$ y $\int\limits_0^\infty 9te^{-3t}dt = 1.0$.

   La vida media se calcula con la expresión:

$$\mu = \int\limits_0^\infty t\left(9te^{-3t}\right)dt = 9\int\limits_0^\infty t^2e^{-3t}dt \tag{9.1}$$

   Realizando la integral del tipo $\int t^n e^{at}dt$ por partes ($u = t^n, dv = e^{at}dt$, $du = nt^{n-1}$, $v = \frac{e^{at}}{a}$) obtenemos la siguiente fórmula de recurrencia:

$$\int t^n e^{at}dt = t^n\frac{e^{at}}{a} - \frac{n}{a}\int t^{n-1}e^{at}dt$$

   Aplicándola sucesivamente a la expresión 9.1, obtenemos $\mu = 9\int\limits_0^\infty t^2e^{-3t}dt = -\frac{2}{3}e^{-3t} - 2te^{-3t} - 3t^2e^{-3t}\Big|_0^\infty = \frac{2}{3}$.

   Si identíficamos la función densidad dada como una Erlang de parámetros $n = 2$, $\lambda = 3$, podemos calcular su vida media como:

$$\mu = \frac{n}{\lambda} = \frac{2}{3}$$

2. $P(\tau > 1) = 9\int\limits_1^\infty te^{-3t}dt = 9\int te^{-3t}dt = 9\left(-\frac{1}{9}e^{-3t} - \frac{1}{3}te^{-3t}\right)\Big|_1^\infty = 0.199\,15$

3. $10000 \times 0.19915 = 1991.\,5$ elementos

4. $P(\tau > 2/\tau > 1) = \dfrac{9\displaystyle\int_2^\infty te^{-3t}dt}{9\displaystyle\int_1^\infty te^{-3t}dt} = \frac{1.7351\times10^{-2}}{0.19915} = 8.7125 \times 10^{-2}$

5. $h(t) = \frac{f(t)}{R(t)}$, $h(1) = \frac{f(1)}{R(1)} = \frac{9\times1\times e^{-3\times1}}{P(\tau>1)} = \frac{0.44808}{0.19915} = 2.2500$ $\text{horas}^{-1} = \frac{2.25}{60}$ $\text{minutos}^{-1} = 0.0375$ minutos^{-1}

6. $P(\tau < 1 + 1/60/\tau > 1) = \dfrac{9\displaystyle\int_1^{1+\frac{1}{60}} te^{-3t}dt}{9\displaystyle\int_1^\infty te^{-3t}dt} = 0.03688$

   Nota complementaria: El modelo teórico nos indica que si sobrevivieran una hora los 1991.5 esperados, en el minuto siguiente se esperan que se rompan $1991.5 \times 0.03688 = 73.447$ elementos.

   Se puede dar una respuesta aproximada usando la tasa de fallo por minuto correspondiente al momento $t = 1$ *hora* : $1991.5 \times 0.0375 = 74.681$ elementos.

   La aproximación sería aún mejor si usaramos los datos en el punto medio del intervalo entre una hora y una hora y 1 minuto, $t = 1 + \frac{1}{60\times2}$

   En efecto:

   $$h(1 + \tfrac{1}{120}) = \frac{f(1+\frac{1}{120})}{R(1+\frac{1}{120})} = \frac{9\times\left(1+\frac{1}{120}\right)\times e^{-3\times\left(1+\frac{1}{120}\right)}}{9\displaystyle\int_{1+\frac{1}{120}}^\infty te^{-3t}dt} = 2.2547$$

   Los que sobrevivirían hasta 1 hora y medio minuto serían

   $$10000 \times 9 \int_{1+\frac{1}{120}}^\infty te^{-3t}dt = 1954.5$$

   Usando estos dos valores obtendríamos.

   $2.2547/60 \times 1954.5 = 3.7578 \times 10^{-2} \times 1954.5 = 73.446$ elementos.

Ejercicio 110 *Suponiendo que la distribución del tiempo de duración sin fallos del disco duro de un ordenador sigue una distribución uniforme en el intervalo de tiempo de 100 horas a 1500 horas se pide:*

1. *Encontrar las funciones de distribución de fiabilidad y tasa de fallo.*

2. *Representar gráficamente la función tasa de fallo en dicho intervalo. ¿Se puede deducir que sufre desgaste este tipo de disco duro?*

3. *¿Cuál es la probabilidad de que un disco de estas características duré 500 horas?*

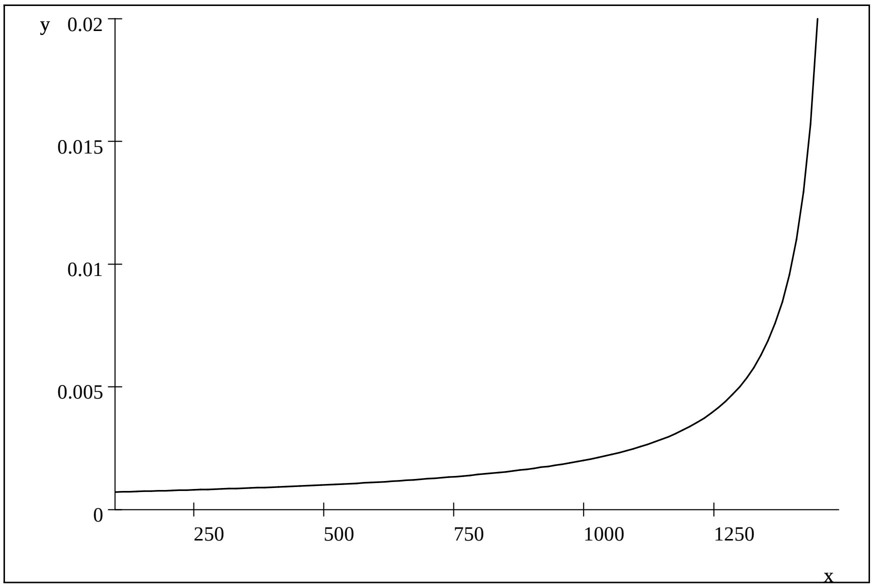
4. *¿Cuál es la probabilidad de que un disco de estas características y que ya haya durado 500 horas, dure todavía 500 horas más ?*

1. $f(t) = \frac{1}{1400}$, para $100 \leq t \leq 1500$, $\quad f(t) = 0$ en el resto

   Considerando únicamente el intervalo $100 \leq t \leq 1500$

   (a) $F(t) = \int_{100}^{t} \frac{1}{1400} dt = \frac{t-100}{1400}$, $R(t) = 1 - F(t) = 1 - \frac{t-100}{1400} = \frac{1500-t}{1400}$
   $h(t) = \frac{\frac{1}{1400}}{\frac{1500-t}{1400}} = \frac{1}{1500-t}$

2. Representamos $h(t) = \frac{1}{1500-t}$



   (a) Sufre desgaste porque la tasa de fallo es creciente.

3. $R(500) = \frac{1500-500}{1400} = 0.\,714\,29$

4. $\frac{F(1000)-F(500)}{R(500)} = \frac{\frac{1500-500}{1400} - \frac{1500-1000}{1400}}{\frac{1500-500}{1400}} = 0.\,5$

Ejercicio 111 *Un componente electrónico tiene una función tasa de fallo constante e igual a 0.005 fallos/hora. Calcular:*

1. *Su función de fiabilidad*
2. *Su vida media*
3. *La probabilidad de que este componente dure más de 125 horas*
4. *Si un componente de este tipo ya ha durado 125 horas, ¿Cúal es la probabilidad de que dure 125 horas más?*

1. La función con tasa de fallo constante es la exponencial, por tanto su función de fiabilidad es $e^{-0.005t}$
2. La vida media es la inversa de la tasa de fallo: $1/0.005 = 200$ horas.
3. $R(125) = e^{-0.005\times 125} = 0.535\,26$
4. $\frac{R(250)}{R(125)} = \frac{e^{-0.005\times 250}}{0.53526} = \frac{0.286\,5}{0.53526} = 0.535\,25$. Permanece el mismo valor, porque en esta función la tasa de fallo es constante (no tiene memoria).

Ejercicio 112 *El tiempo en horas que la batería de una calculadora mantiene su carga es una variable aleatoria T. Suponemos que esta variable aleatoria sigue una distribución cuya función densidad es* $f(t) = 0.02\ t\ e^{-0.01t^2}$

1. *¿Cuál es la función de fiabilidad? ¿Cuál es la fiabilidad para $t = 12$ horas?*
2. *¿Cuál es la probabilidad de que la batería dure al menos 3 horas?*
3. *Calcular la función tasa de fallo, indicando si es una función creciente o decreciente*

1. $R(t) = 1 - F(t) = 1 - \int_0^t 0.02te^{-0.01t^2}dt = \exp\left(-.0\,01\ t^2\right)$
   $R(12) = \exp\left(-.0\,1 \times 12^2\right) = 0.\,236\,93$
2. $R(3) = \exp\left(-.0\,1 \times 3^2\right) = 0.\,913\,93$
3. $h(x) = \frac{f(t)}{R(t)} = \frac{0.02te^{-0.01t^2}}{\exp(-0.\,01t^2)} = 0.02\ t$
   Por tanto la tasa de fallo es creciente.La batería sufre desgaste.

Ejercicio 113 *Suponiendo que una distribución de tiempo de fallo esta dado por una distribución uniforme:*

$$\begin{cases} f(t) = \frac{1}{5} & si\ 0 \leq t \leq 5 \\ f(t) = 0 & en\ el\ resto \end{cases}$$

1. *Determinar la función de infiabilidad*

2. *Determinar la función de fiabilidad*

3. *Calcular la probabilidad de que las unidades que se ajusten a esta distribución duren entre 3 y 4 horas*

4. *Determinar la función tasa de fallo e indicar si el modelo sería adecuado para piezas que sufran desgaste.*

1. $F(t) = \int_{-\infty}^{t} f(t)dt$ Por tanto en este caso:

$$\begin{cases} F(t) = 0 & si \quad t \leq 0 \\ F(t) = 0 + \int_0^t \frac{1}{5}dt = \frac{1}{5}t & si \ 0 \leq t \leq 5 \\ F(t) = 1 & si \quad t \geq 5 \end{cases}$$

La función de fiabilidad se obtiene restando de 1 la anterior función, por tanto

$$\begin{cases} R(t) = 1 & si \quad t \leq 0 \\ R(t) = 1 - \int_0^t \frac{1}{5}dt = 1 - \frac{1}{5}t & si \ 0 \leq t \leq 5 \\ R(t) = 0 & si \quad t \geq 5 \end{cases}$$

$P(3 < t < 4) = F(4) - F(3) = \frac{4}{5} - \frac{3}{5} = 0.2$

La tasa de fallo es

$$\begin{cases} h(t) = \frac{f(t)}{R(t)} = \frac{0}{1} = 0 & si \quad t \leq 0 \\ h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{5}}{1-\frac{1}{5}t} = \frac{1}{5-t} & si \ 0 \leq t \leq 5 \\ h(t) = \frac{f(t)}{R(t)} = \frac{0}{0} & si \quad t \geq 5 \end{cases}$$

En el último intervalo la tasa de fallo está indeterminada. Ya que ningun elemento sobrevive despues de $t = 5$, sería también razonable definir en este intervalo el valor 0 para la tasa de fallo.

La tasa de fallo es creciente en el periodo de vida [0,5] por lo que podría aplicarse a elementos que sufran desgaste.

Ejercicio 114 *A partir de los datos de la siguiente tabla, correspondiente al momento en que han fallado ciertos dispositivos, calcular un valor aproximado para la tasa de fallo por minuto, correspondiente a cada intervalo*

Intervalo de tiempo	*1ª hora*	*2ª hora*	*3ª hora*	*4ª hora*
nº de fallos en el intervalo	*30*	*20*	*15*	*10*
Elementos supervivientes al principio del intervalo	*500*	*470*	*450*	*435*

Int. Tmp = Intervalo de tiempo
Est. sup. = Estimación de elementos supervivientes a la mitad del periodo
t. fallo = Tasa de fallo por minuto

Int. Tmp	1^a hora	2^a hora	3^a hora	4^a hora
Est. sup.	485	460	442.5	430
t. fallo	$\frac{30}{485}/60 = 1.03 \times 10^{-3}$	$\frac{20}{460}/60 = 7.25 \times 10^{-4}$	$\frac{15}{442.5}/60 = 5.65 \times 10^{-4}$	$\frac{10}{430}/60 = 3.88 \times 10^{-4}$

T. 10

Distribuciones de tiempos de fallos

Ejercicio 115 *La función de densidad del tiempo de vida de un componente es exponencial:*

$$f(t) = 0.5e^{-0.5t},\ t \geq 0, (t \text{ en meses})$$

1. *Calcula la vida media del componente,así como la tasa de fallo a los 3 meses.*

2. *Si ponemos en funcionamiento un lote de 1000 componentes simultaneamente, ¿Cúantos de estos se espera sobrervivan más de dos meses?*

3. *Cálcula y representa gráficamente las funciones de densidad, ,fiabilidad y tasa de fallo de esta distribución.*

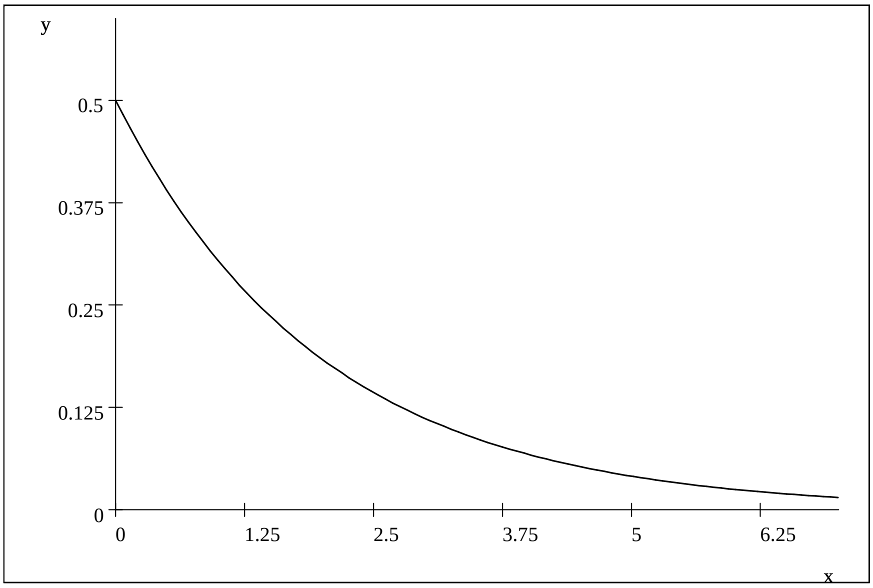
1. La vida media es la inversa de la tasa de fallo en el caso de la distribución exponencial:

$$\mu = \frac{1}{\lambda} = \frac{1}{0.5} = 2meses$$

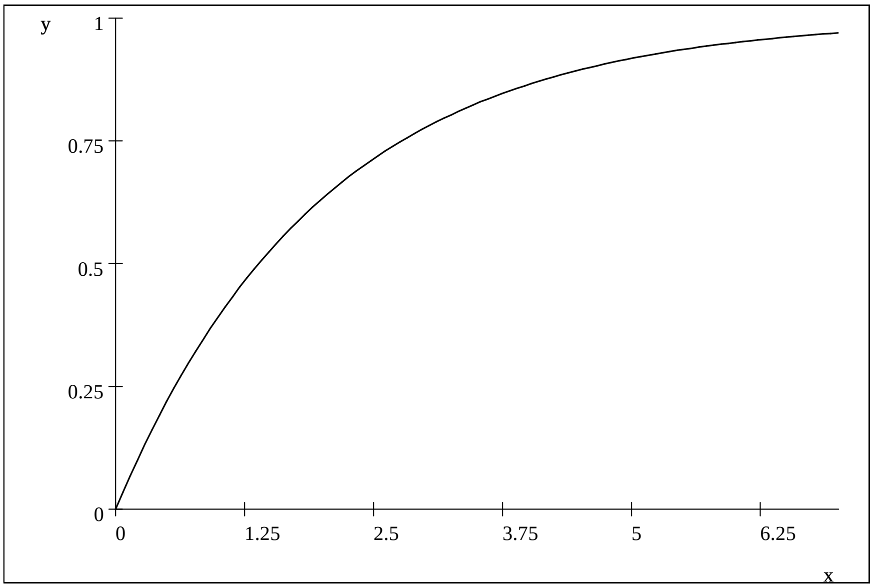
   La tasa de fallo es constante porque la distribución es exponencial. Su valor es siempre 0.5.

2. $1000 \times R(2) = 1000 \times e^{-0.5 \times 2} = 367.88$. Aproximadamente 368 sobrevivirían.

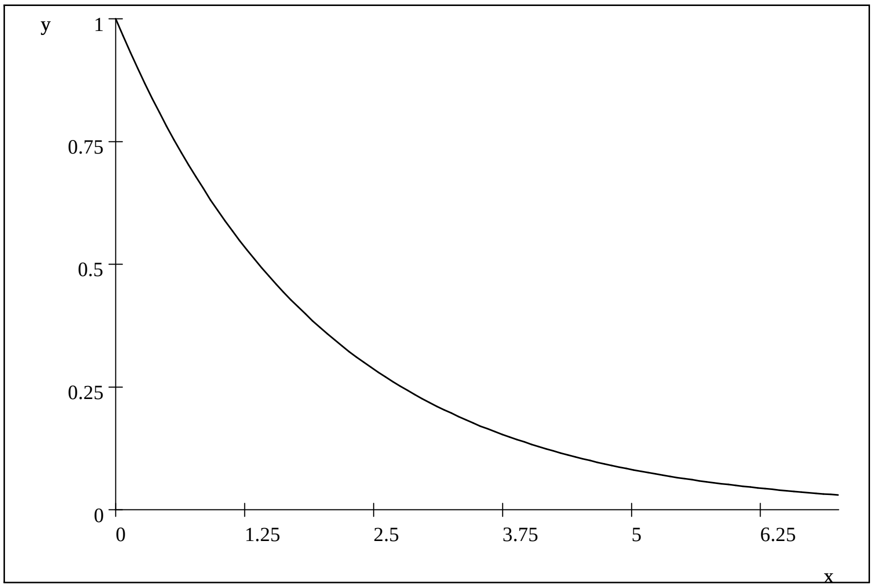
3. Función de densidad: $f(t) = 0.5e^{-0.5t}$



Función de distribución: $F(t) = 1 - e^{-0.5t}$



Función de Fiabilidad: $R(t) = e^{-0.5t}$



Ejercicio 116 *El nº de kilometros recorridos por un modelo de automovil antes que los parachoques resulten inservibles sigue un modelo de distribución lognormal. Se ha observado que el 5% de los parachoques fallan antes de que el vehículo haya recorrido 120000 km y que otro 5% falla despues de que el vehículo haya recorrido más de 180000 km.*

1. *Estimar la media y la desviación típica de la distribución lognormal.*
2. *Hallar el valor de la tasa de fallo a los 150000 km.*
3. *Si se han fabricado 9000 unidades de este automovil. ¿ Cuántos de ellos tendrán los parachoques rotos cuando hayan recorrido 150000 km.?*

1. La expresión de la función de densidad de la distribución lognormal es:

$$f(t) = \frac{1}{t\sigma'\sqrt{2\pi}} e^{-\frac{(\ln t - \mu')^2}{2\sigma'^2}}$$

siendo μ' y σ' la media y la desviación típica de la normal asociada. En este caso t representará el número de km. recorridos por el vehículo cuando se produce el fallo.

De los datos del problema podemos dar valores aproximados para las siguientes probabilidades

$P(t < 120000) = 0.05 \implies P(\ln t < \ln 120000) = 0.05$

$P(t > 180000) = 0.05 \implies P(\ln t > \ln 180000) = 0.05$

Como $\ln t$ se distribuye como una normal de media μ' y desviación típica σ'

Obtenemos

$P(\frac{\ln t-\mu'}{\sigma'} < \frac{\ln 120000-\mu'}{\sigma'}) = 0.05 \implies \frac{\ln 120000-\mu'}{\sigma'} = -1.6449$

$P(\frac{\ln t-\mu'}{\sigma'} > \frac{\ln 180000-\mu'}{\sigma'}) = 0.05 \implies \frac{\ln 180000-\mu'}{\sigma'} = 1.6449$

Resolviendo este sistema obtenemos

$\frac{\ln 120000-\mu'}{\sigma'} - \frac{\ln 180000-\mu'}{\sigma'} = \frac{\ln 2-\ln 3}{\sigma'} = -3.2898 \implies \sigma' = 0.123\,25$

$\frac{\ln 120000-\mu'}{0.123\,25} = -1.6449 \implies \mu' = 11.898$

$\ln 147000 = \ln 147\,000 = 11.898$

Por tanto las estimaciones de la media y la desviación típica de la lognormal son:

$$\hat{\mu} = \exp\left(11.898 + \frac{0.123\,25^2}{2}\right) = 148\ 090\ Km.$$

$$\hat{\sigma} = \sqrt{(\exp(0.123\,25^2) - 1)\ \exp\left(2 \times 11.898 + 0.123\,25^2\right)} = 18322$$

2. La tasa de fallo a los 150000 Km es

$h(150000) = \frac{f(150000)}{R(150000)} =$

$$= \frac{1}{150000 \times 0.123\,25\sqrt{2\pi}} \frac{e^{-\frac{(\ln 150000-11.898)^2}{2\times 0.123\,25^2}}}{P\left(z > \frac{\ln 150000 - 11.898}{0.123\,25}\right)} =$$

$$= \frac{1}{46341.} \frac{e^{-1.368\,5\times 10^{-2}}}{P(z>0.165\,44)} = \frac{1}{46341} \frac{.986\,41}{0.4343} = 4.901\,2 \times 10^{-5}$$

3. $P(t < 150000) = F(150000) = 1 - R(150000) = 1 - 0.4343 = 0.565\,7$

$9000 \times 0.5657 = 5091.3.$

Aproximadamente 5091 vehículos tendrán los parachoques rotos.

Ejercicio 117 *La fiabilidad de los alternadores de unos automóviles es:*

$$R(x) = e^{-\left(\frac{x}{180000}\right)^3}$$

El número de vehículos en que se ha instalado estos alternadores es de 100000. La variable x es el número de kilómetros recorridos antes de la avería.

Se pide:

1. *¿Cuántos alternadores se puede esperar que tengan averías antes de que hayan recorrido 60000Km.*

2. *Calcula la tasa de fallo a los 60000 km y a los 120000 km*

3. *Calcular la vida media de estos alternadores.*

4. *Si la garantia cubre las averías producidas en los primeros 15000 km. ¿Cuántos alternadores puede esperarse que habrá que reparar en garantía?*

1. $1 - R(60000) = 1 - e^{-\left(\frac{60000}{180000}\right)^3} = 3.63596 \times 10^{-2}$

   $0.03635896 \times 100000 = 3636\,alternadores$

2. $h(x) = \frac{f(x)}{R(x)} = \frac{\frac{d}{dx}\left(1 - e^{-\left(\frac{x}{180000}\right)^3}\right)}{e^{-\left(\frac{x}{180000}\right)^3}} = \frac{5.14403 \times 10^{-16} x^2 \exp\left(-1.71468 \times 10^{-16} x^3\right)}{\exp(-1.71468 \times 10^{-16} x^3)} =$
   $5.14403 \times 10^{-16} x^2$

   $h(60000) = 5.14403 \times 10^{-16} \times 60000^2 = 1.85185 \times 10^{-6}$

   $h(120000) = 5.14403 \times 10^{-16} \times 120000^2 = 7.4074 \times 10^{-6}$

3. La media de la distribución de Weibul es

   $\mu = \gamma + \eta\Gamma\left(1 + \frac{1}{\beta}\right) = 0 + 180000 \times \Gamma\left(1 + \frac{1}{3}\right) = 180000 \times 0.894 = 160925Km.$

   El valor de $\Gamma\left(1 + \frac{1}{3}\right)$ suele venir en el papel de Weibul. La función gamma se define como

   $\Gamma(y) = \int_0^\infty e^{-x} x^{y-1} dx \quad \Longrightarrow \quad \Gamma\left(1 + \frac{1}{3}\right) = \int_0^\infty e^{-x} x^{\frac{1}{3}} dx$

4. $1 - R(15000) = 1 - e^{-\left(\frac{15000}{180000}\right)^3} = 1 - e^{-\frac{1}{1728}} = 5.78536 \times 10^{-4}$

   $5.78536 \times 10^{-4} \times 100000 = 57.8536$. 58 alternadores aproximadamente.

Ejercicio 118 *Cien unidades se han sometido a una prueba de vida hasta que han fallado. Se han obtenido los datos siguientes:*
t = tiempo en horas de duración de las unidades,
n = Número de unidades que han durado este tiempo

t	*0-100*	*100-200*	*200-300*	*300-400*	*400-500*	*Más de 500*
n	*50*	*18*	*17*	*8*	*4*	*3*

¿Se puede admitir que una distribución exponencial con valor medio 160 horas representa razonablemente los tiempo de fallo del modelo del que proceden estos datos?

El Parámetro de la distribución exponencial es: $\lambda = \frac{1}{\mu} = \frac{1}{160}$

La función de densidad es por tanto: $f(t) = \frac{1}{160}e^{-\frac{1}{160}t}$ *para* $t \geq 0$ y la función de distribución es $F(t) = \int_0^t \frac{1}{160}e^{-\frac{1}{160}t}dt = 1 - e^{-\frac{1}{160}t}$

Calculamos ahora los valores teóricos que correspondería a la probabilidad de los distintos sucesos de la tabla:

$P_1 = P(0 < t \leq 100) = \left(1 - e^{-\frac{1}{160}100}\right) - \left(1 - e^{-\frac{1}{160}0}\right) = 0.46474$

$P_2 = P(100 < t \leq 200) = \left(1 - e^{-\frac{1}{160}200}\right) - \left(1 - e^{-\frac{1}{160}100}\right) = 0.24876$

$P_3 = P(200 < t \leq 300) = \left(1 - e^{-\frac{1}{160}300}\right) - \left(1 - e^{-\frac{1}{160}200}\right) = 0.13315$

$P_4 = P(300 < t \leq 400) = \left(1 - e^{-\frac{1}{160}400}\right) - \left(1 - e^{-\frac{1}{160}300}\right) = 0.07127$

$P_5 = P(400 < t \leq 500) = \left(1 - e^{-\frac{1}{160}500}\right) - \left(1 - e^{-\frac{1}{160}400}\right) = 3.8148 \times 10^{-2}$

$P_6 = P(500 < t) = 1 - \left(1 - e^{-\frac{1}{160}500}\right) = 4.3937 \times 10^{-2}$

En la siguiente tabla se muestra las frecuencias experimentales y las esperadas, siendo $n = 100$, el número total de elementos de la muestra

t	0-100	100-200	200-300	300-400	400-500	Más de 500
n	50	18	17	8	4	3
nP$_i$	46.474	24.876	13.315	7.127	3.8148	4.3937

Debido a que las dos últimas casillas contienen menos de 5 elementos las agrupamos en una única clase obteniendose

t	0-100	100-200	200-300	300-400	Más de 400
n	50	18	17	8	7
nP$_i$	46.474	24.876	13.315	7.127	8.2085

Calculamos el valor de la chi-cuadrado experimental. Los grados de libertad que corresponden a este valor de chi-cuadrado es $k-p-1$, siendo $k = 5$ el número de clases y $p = 0$ el número de parámetros estimados con la muestra.

$\chi^2_{\exp} = \sum_{i=1}^5 \frac{(n_i - nP_i)^2}{nP_i} = \frac{(50-46.474)^2}{46.474} + \frac{(18-24.876)^2}{24.876} + \frac{(17-13.315)^2}{13.315} + \frac{(8-7.127)^2}{7.127} + \frac{(7-8.2085)^2}{8.2085} = 3.4728$

Comparando con el valor de la chi-cuadrado teórica al 5% de nivel de significación se tiene que

ChiSquareInv$(.95; 4) = 9.4877 > 3.4728$

Por lo tanto no se puede rechazar la hipótesis nula de que la muestra siga una distribución exponencial de media 160, así que admitiremos dicha hipótesis.

Ejercicio 119 *Los tiempos, en horas, de duración en funcionamiento de 20 baterias ha sido:*

26, 32, 34, 39, 56, 71, 84, 88, 89, 95, 98, 113, 118, 119, 123, 127, 160, 219, 224, 242

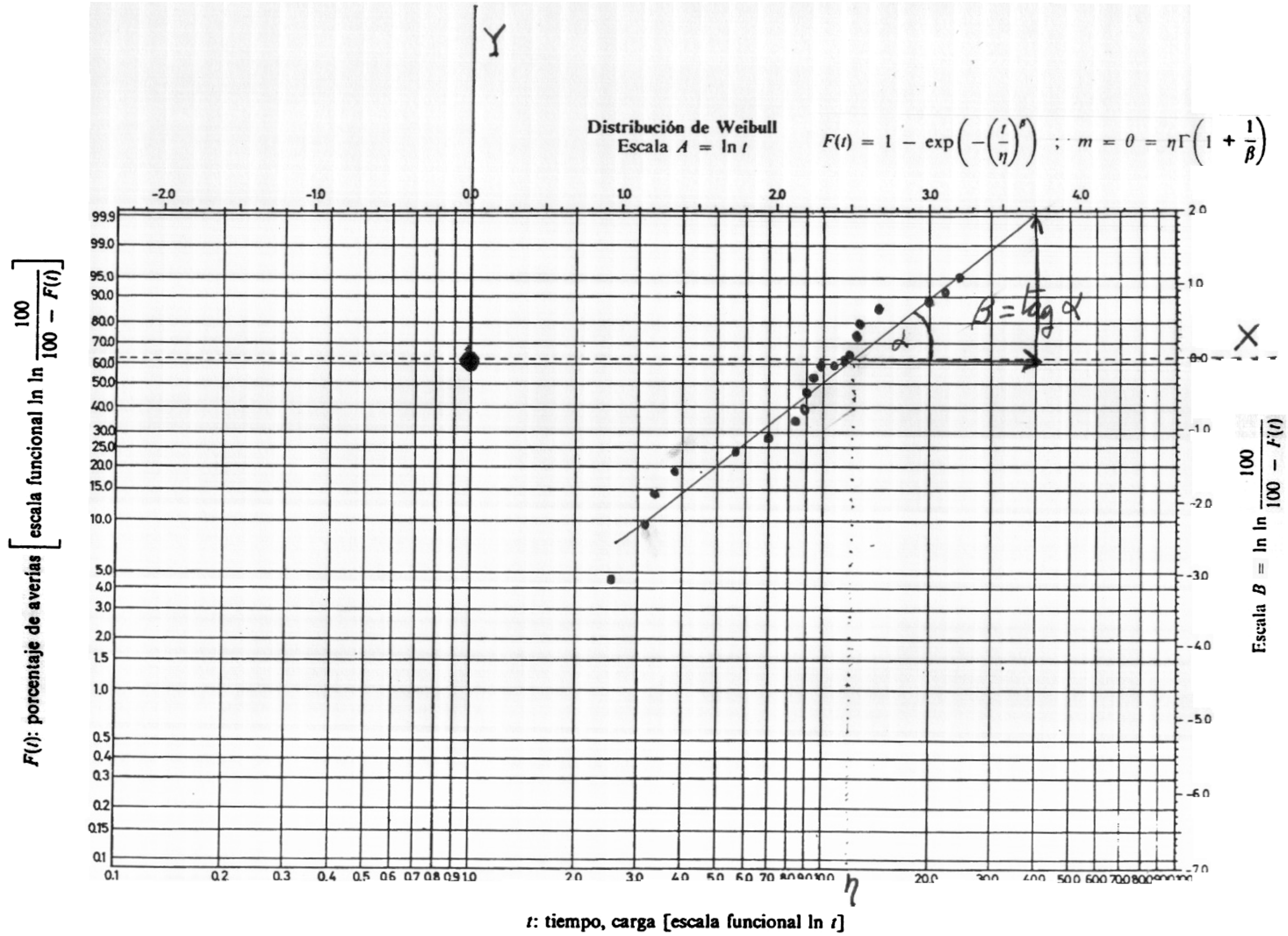
Ajustar una distribución de Weibull a estos datos usando los procedimientos siguientes:

1. *Por medio de papel de Weibull.*
2. *Numéricamente, realizando un ajuste de regresión lineal*
3. *Por medio del procedimiento* ***Distribution Fitting*** *de Statgraphics.*

1. Los valores de $(t,\ F(t))$ aparecen en las tablas del apartado b). La representación de los datos sobre papel de Weibull puede verse en la gráfica de la página 122. Se han representado los valores de t divididos por 10 para que entren todos los valores en la gráfica y además el dibujo quede centrado. El origen de las coordenadas $X,\ Y$, es el círculo negro más grande. En esta gráfica se observa que la pendiente vale aproximadamente $\frac{2}{1.2} = 1.666\,7$. que es el valor que asignaremos a β. El valor de η se estima con el valor de t en la intersección de la recta con el eje de las X. El valor correspondiente está entre 100 y 200. Para dar un valor concreto adoptaremos 130 como valor de η.

   Si queremos estimar la vida media , $\mu = \gamma + \eta\Gamma\left(1+\frac{1}{\beta}\right)$, usamos la tabla de la derecha. Tomando $\gamma = 0$, $\Gamma\left(1+\frac{1}{\beta}\right) = \frac{\mu}{\eta}$ ($\frac{m}{n}$ en la tabla). En este caso para $\beta = 1.6667$ (tomamos el más cercano 1.7) el valor que corresponde es 0.892.

   Por tanto la vida media es $\mu = \gamma + \eta\Gamma\left(1+\frac{1}{\beta}\right) = 0 + 130 \times 0.892 = 115.\,96$ horas.



β	$\frac{m}{\eta}$
0,2	12,0
0,3	9,261
0,4	3,323
0,5	2,000
0,6	1,505
0,7	1,266
0,8	1,133
0,9	1,052
1,0	1,000
1,1	0,965
1,2	0,941
1,3	0,923
1,4	0,911
1,5	0,903
1,6	0,897
1,7	0,892
1,8	0,889
1,9	0,887
2,0	0,886
2,1	0,886
2,2	0,886
2,3	0,886
2,4	0,886
2,5	0,887
2,6	0,888
2,7	0,889
2,8	0,890
2,9	0,892
3,0	0,894
3,1	0,895
3,2	0,896
3,3	0,897
3,4	0,898
3,5	0.900
3,6	0.901
3,7	0.902
3,8	0,904
3,9	0,905
4,0	0.906

2. Para ello hacemos los cambios que permitan representar la función de Weibull en una recta.

t	26	32	34	39	56
$F(t)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$
$\ln t$	$3.258\,1$	$3.465\,7$	$3.526\,4$	$3.663\,6$	$4.025\,4$
$\ln\ln\left(\frac{1}{1-F(t)}\right)$	$-3.020\,2$	$-2.301\,8$	$-1.869\,8$	$-1.554\,4$	$-1.302\,2$

t	71	84	88	89	95
$F(t)$	$\frac{6}{21}$	$\frac{7}{21}$	$\frac{8}{21}$	$\frac{9}{21}$	$\frac{10}{21}$
$\ln t$	$4.262\,7$	$4.430\,8$	$4.477\,3$	$4.488\,6$	$4.553\,9$
$\ln\ln\left(\frac{1}{1-F(t)}\right)$	$-1.089\,2$	$-.902\,72$	$-.734\,86$	$-.580\,5$	$-.435\,99$

t	98	113	118	119
$F(t)$	$\frac{11}{21}$	$\frac{12}{21}$	$\frac{13}{21}$	$\frac{14}{21}$
$\ln t$	4.585	$4.727\,4$	$4.770\,7$	$4.779\,1$
$\ln\ln\left(\frac{1}{1-F(t)}\right)$	$-.298\,49$	$-.165\,7$	$-3.554\,3\times 10^{-2}$	$9.404\,8\times 10^{-2}$

t	123	127	160	219	224	242
$F(t)$	$\frac{15}{21}$	$\frac{16}{21}$	$\frac{17}{21}$	$\frac{18}{21}$	$\frac{19}{21}$	$\frac{20}{21}$
$\ln t$	$4.812\,2$	$4.844\,2$	$5.075\,2$	$5.389\,1$	$5.411\,6$	$5.488\,9$
$\ln\ln\left(\frac{1}{1-F(t)}\right)$	0.2256	0.3612	0.5058	0.6657	0.855	1.1133

Considerando $X = \ln t$, $Y = \ln\ln\left(\frac{1}{1-F(t)}\right)$, obtenemos la recta de regresión correspondiente a los valores de las dos últimas filas. La ecuación de esta recta es:

$$y - \overline{Y} = \frac{S_{XY}}{S_X^2}(x - \overline{X})$$

siendo $\bar{X}$, $\overline{Y}$, S_{XY}, S_X^2 respectivamente la media de X, la media de Y, la covarianza de X, Y, y la varianza de X.

La ecuación de la recta resulta:

$Y = 1.66X - 7.995$

$\beta = 1.66$

$-\beta \ln\eta = -7.995 \Longrightarrow -1.66\ln\eta = -7.995 \Longrightarrow \ln\eta = \frac{-7.995}{-1.66} = 4.816\,3$; $\quad \eta = \exp(4.816\,3) = 123.51$

3. Se introducen los datos de tiempos de fallo (t) en una columna del fichero de datos. Seleccionamos

Describe $\rightarrow$ **Distribution** $\rightarrow$

Distribution Fitting (**Uncensored Data**)

y obtenemos el siguiente resultado:

Uncensored Data - Col_1

Analysis Summary

Data variable: Col_1

20 values ranging from 26.0 to 242.0

Fitted Weibull distribution:

shape = 1.85365

scale = 121.89

Ejercicio 120 *Los elementos fabricados por un cierto proceso tienen una duración (en meses) cuya función tasa de fallo viene dada por* $h(t) = t^2$ *para* $t > 0$ *Hallar:*

1. *la función de densidad, de fiabilidad y de infiabilidad*
2. *La probabilidad de que uno de estos dispositivos dure más de 1 mes*
3. *La producción de un mes es de 10000 elementos. Si hay que reponer todos los dispositivos que duren menos de medio mes. ¿ Cuántos elementos se puede esperar que haya que reponer de estos 10000?*
4. *Da una expresión para la vida media de estos dispositivos*

1. $R(t) = e^{-\int_0^t t^2 dt} = e^{-\frac{1}{3}t^3}$

   $F(t) = 1 - R(t) = 1 - e^{-\frac{1}{3}t^3}$

   $f(t) = F'(t) = \frac{d}{dt}(1 - e^{-\frac{1}{3}t^3}) = t^2 e^{-\frac{1}{3}t^3}$

2. $R(1) = e^{-\frac{1}{3}1^3} = 0.716\,53$

3. $F(t) = 1 - R(t) = 1 - e^{-\frac{1}{3}t^3}$; $F(0.5) = 1 - e^{-\frac{1}{3}0.5^3} = 4.081\,1 \times 10^{-2}$

   $10000 \times 4.081\,1 \times 10^{-2} = 408.11$ Aproximadamente 408 unidades habrá que reponer

4. $\int_0^\infty e^{-\frac{1}{3}t^3} dt$ o también $\int_0^\infty t(t^2 e^{-\frac{1}{3}t^3}) dt$.

   Teniendo en cuenta que la distribución es de Weibul el valor de la vida media podría hallarse por la expresión:

Ejercicio 121 *El tiempo de fallo (en horas) de un dispositivo sigue una distribución de probabilidad cuya función de densidad es* $f(t) = \alpha^2 te^{-\alpha t}$, $t > 0$, $\alpha > 0$.

1. *Calcular, en función de* α *la probabilidad de que un componente que ya haya durado 100 horas dure 100 horas más.*

2. *El coste de producir un componente es proporcional al cuadrado de su vida media* ($k\mu^2$) *y se estima que la ganancia obtenida por cada uno de estos componentes es de 48 euros por cada hora que funciona sin fallar. Calcular una expresión en función de* α *y* k *para el beneficio medio obtenido con estos componentes.*

3. *Demostrar que el beneficio máximo que se obtiene corresponde a un valor de* $\alpha = \frac{K}{12}$ *con un valor medio de* $\frac{576}{K}$ *euros por unidad.*

1. La integral indefinida de la función densidad que se realiza por partes tomando $u = t,\ dv = e^{-\alpha t}$

   es $\int \alpha^2 te^{-\alpha t}dt = \alpha^2 \int te^{-\alpha t}dt = \alpha^2 \left(-\frac{te^{-\alpha t}}{\alpha} - \frac{e^{-\alpha t}}{\alpha^2}\right) = -\alpha e^{-\alpha t}t - e^{-\alpha t}$

   $F(t) = \int_0^t \alpha^2 te^{-\alpha t}dt = \left(-\alpha e^{-\alpha t}t - e^{-\alpha t}\right)\Big|_0^t = -\alpha te^{-\alpha t} - e^{-\alpha t} + 1$

   $P(\tau \geq 200/\tau \geq 100) = \frac{P(\tau \geq 200)}{P(\tau \geq 100)} = \frac{\alpha \times 200 \times e^{-\alpha 200} + e^{-\alpha \times 200}}{\alpha \times 100 \times e^{-\alpha 100} + e^{-\alpha \times 100}} = e^{-100\alpha}\frac{200\alpha+1}{100\alpha+1}$

   $\mu = \int_0^\infty \left(\alpha te^{-\alpha t} + e^{-\alpha t}\right)dt = \left(-e^{-\alpha t}t - 2e^{-\alpha t}/\alpha\right)\Big|_0^\infty = \frac{2}{\alpha}$

2. $Beneficio\, medio = 48 \times \frac{2}{\alpha} - k \times 4/\alpha^2$

3. $\frac{d(48 \times \frac{2}{\alpha} - 4k/\alpha^2)}{d\alpha} = -8\frac{12\alpha - k}{\alpha^3} = 0; \quad \alpha = \frac{k}{12}$

$$Beneficio\, máximo = 48 \times \frac{2}{\alpha} - 4k/\alpha^2 = 48 \times \frac{2}{\frac{k}{12}} - 4k/\left(\frac{k}{12}\right)^2 = \frac{576}{k}$$

Ejercicio 122 *Un submarinista, que ha de reparar una plataforma petrolífica, puede elegir entre dos equipos de buceo. La reparación se realiza en condiciones peligrosas y el equipo de buceo puede fallar. La distribución del tiempo de fallo de los estos equipos sigue una distribución de Weibull de parámetros a)* $\eta = 1$, $\beta = 2$ *b)* $\eta = 2$, $\beta = 1$

1. *¿Qué equipo debe usar si la reparación dura una hora?*

2. *Responder a la misma pregunta si ambos equipos han sido ya usados, sin fallos, durante 3.17 horas.*

1. Calculamos la probabilidad de que cada uno de estos equipos dure al menos una hora.
   $R_a(t) = e^{-\left(\frac{t}{1}\right)^2}$; $R_a(1) = R_a(t) = e^{-\left(\frac{1}{1}\right)^2} = 0.367\,88$
   $R_b(t) = e^{-\left(\frac{t}{2}\right)^1}$; $R_b(1) = R_b(t) = e^{-\left(\frac{1}{2}\right)^1} = 0.606\,53$
   Es más recomendable el segundo aunque, si las condiciones son peligrosas, es también poco recomendable.

2. Calculamos la probabilidad de que cada uno de estos equipos dure, al menos, una hora más si ya han durado 3.17 horas.
   $P_a(t > 4.17/t > 3.17) = \frac{e^{-\left(\frac{4.17}{1}\right)^2}}{e^{-\left(\frac{3.17}{1}\right)^2}} = 6.\,490\,5 \times 10^{-4}$
   $P_b(t > 4.17/t > 3.17) = \frac{e^{-\left(\frac{4.17}{2}\right)^1}}{e^{-\left(\frac{3.17}{2}\right)^1}} = 0.606\,53$
   Es mucho más recomendable el segundo. Como la distribución del segundo es exponencial, este equipo no sufre desgaste. En cambio el primero ha sufrido mucho desgaste.

T. 11

Modelos para Sistemas. Redundancia

Ejercicio 123 *La duración (en horas) de unos dispositivos se rige por una distribución cuya función densidad viene dada por:*

$$\begin{array}{cl} f(t) = \frac{1}{8}t & si\ t \in [0,\ 4] \\ f(t) = 0 & en\ el\ resto \end{array}$$

1. *Calcular la función de Infiabilidad, la función de Fiabilidad y la vida media.*

2. *Calcular la probabilidad de que uno de estos dispositivos dure mas de 3 horas*

3. *Probabilidad de que un sistema formado por dos de estos dispositivos en paralelo dure más de 3 horas*

1. $F(t) = \int_0^t \frac{1}{8}tdt = \frac{1}{16}t^2 \;\; si\ t \in [0,\ 4]$

   $R(t) = 1 - \frac{1}{16}t^2 \quad si\ t \in [0,\ 4]$

   Para calcular la vida media usamos dos procedimientos:

   $\mu = \int_0^4 t\frac{1}{8}tdt = 2.6667$

   $\mu = \int_0^4 \left(1 - \frac{1}{16}t^2\right) dt = 2.6667$

2. $1 - \frac{1}{16}3^2 = 0.4375$

3. $1 - (1 - 0.4375)^2 = 0.68359$

Ejercicio 124 *Tres componentes con tiempo de fallo exponencial y tasa de fallo 0.03, 0.06 y 0.04 se han dispuesto formando un sistema en serie.*

1. *Hallar R(6) para el sistema*

2. *Calcular la vida media del sistema*

3. *Calcular la probabilidad de que el sistema permanezca funcionando al menos 4 horas*

1. La fiabilidad del sistema es el producto de las fiabilidades de sus componentes resultando otra exponencial cuya tasa de fallos es la suma de la tasa de fallo de sus componentes $R_s(t) = e^{-(0.03+0.06+0.04)t} = e^{-0.\,13t}$.

   $R(6) = e^{-.\,13\times 6} = .\,458\,40.$

   La media de una exponencial es $\frac{1}{\lambda} = \frac{1}{0.13} = 7.\,692\,3$

   $R_s(4) = e^{-.\,13\times 4} = 0.594\,52 = 0.594\,52$

Ejercicio 125 *La fiabilidad (tiempo hasta el fallo en horas) de un dispositivo viene dada por la función:* $R(t) = e^{-\frac{t}{10}}$

1. *Calcular la función de densidad, la tasa de fallos y la vida media de estos dispositivos.*

2. *Calcular la probabilidad de que uno de estos dispositivos dure más de 9 horas*

3. *Para aumentar la fiabilidad del sistema se han colocado 4 de estos dispositivos en paralelo. ¿ Cuál es la probabilidad de que este sistema duré más de 9 horas?*

4. *Calcular la vida media de este sistema*

1. $f(t) = -R'(t) = -\frac{d}{dt}\left(e^{-\frac{t}{10}}\right) = \frac{1}{10}e^{-\frac{1}{10}t}$

   $h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{10}e^{-\frac{1}{10}t}}{e^{-\frac{t}{10}}} = \frac{1}{10}$

   La media de la distribución exponencial es la inversa del parámetro λ. En este caso 10 horas

2. $R(9) = e^{-\frac{9}{10}} = 0.\,406\,57$

3. $1 - (1 - 0.\,406\,57)^4 = 0.\,875\,984$

4. $\mu_s = \mu + \frac{\mu}{2} + \frac{\mu}{3} + \frac{\mu}{4} = 10 + 5 + 3.33 + 2.5 = 20.\,83$ *horas*

Ejercicio 126 *Un sistema con tres componentes independientes trabaja correctamente si al menos uno de ellos funciona. Las tasas de fallo de cada uno de ellos son: 0.01, 0.02, 0.03. Suponiendo que el tiempo de vida de estos componentes sigue una distribución exponencial, calcular:*

1. *La función de Fiabilidad del sistema*

2. *La probabilidad de que el sistema funcione al menos 100 horas.*

3. *La tasa de fallo del sistema*

1. La función de fiabilidad de cada componente es $R(t) = \exp\left(-\int_0^t kdt\right) = \exp(-kt)$.

   La Probabilidad de que uno de ellos no funcione más de t es $1 - R(t) = 1 - \exp(-kt)$

   La probabilidad de que ninguno funcione más de t es:

   $(1 - \exp(-0.01t))(1 - \exp(-0.02t))(1 - \exp(-0.03t))$. Por tanto la probabilidad de que al menos uno funcione despues de$t$, es la fiabilidad del sistema:

   $R_s(t) = 1 - (1 - \exp(-0.01t))(1 - \exp(-0.02t))(1 - \exp(-0.03t))$

2. La probabilidad de que el sistema funcione al menos 100 horas es

   $R_s(100) =$

   $= 1 - (1 - \exp(-1.0))(1 - \exp(-2.0))(1 - \exp(-3.0)) = 0.48064$

3. La tasa de fallo del sistema es

   $h(t) = \frac{f(t)}{R(t)} = \frac{-\frac{d}{dt}[1-(1-\exp(-0.01t))(1-\exp(-0.02t))(1-\exp(-0.03t))]}{1-(1-\exp(-0.01t))(1-\exp(-0.02t))(1-\exp(-0.03t))} =$

   $= \frac{-\frac{d}{dt}(\exp(-.02t)-1.0\exp(-.05t)+\exp(-.01t)-1.0\exp(-.04t)+\exp(-.06t))}{\exp(-.02t)-1.0\exp(-.05t)+\exp(-.01t)-1.0\exp(-.04t)+\exp(-.06t)} =$

$$= \frac{.02e^{-.02t} - .05e^{-.05t} + .01e^{-.01t} - .04e^{-.04t} + .06e^{-.06t}}{e^{-.02t} - 1.0e^{-.05t} + e^{-.01t} - 1.0e^{-.04t} + e^{-.06t}}$$

Ejercicio 127 *Se tienen tres componentes A, B, C en serie con fiabilidad 0.5, 0.8, 0.85. Se desea mejorar la fiabilidad del sistema añadiendo redundancia activa componente a componente.*

1. *Con un solo elemento*

2. *Con dos elementos*

3. *Con tres elementos*

¿Cuál es la mejor composición del sistema en cada caso?

Los elementos separados por comas están en serie. Los consecutivos en paralelo.

1. $R(A,B,C) = 0.5 \times 0.8 \times 0.85 = 0.34$

2. Añadiendo un solo elemento

   $R(AA,B,C) = (1-(1-0.5)^2) \times 0.8 \times 0.85 \quad = 0.51$. Este caso es el mejor añadiendo un solo elemento, ya que

   $R(A,BB,C) = 0.5 \times (1-(1-0.8)^2) \times .85 < 0.5$

   $R(A,B,CC) < 0.5$

3. Con dos elementos. Los sistemas se presentan a continuación ordenados por el número de elementos en redundancia con A.

   $R(AAA,B,C) = (1-(1-.5)^3) \times 0.8 \times 0.85 = .875 * 0.8 * 0.850 = 0.595$

   $R(AA,BB,C) = 0.75 \times 0.96 \times 0.85 = .612$ Este es el mejor con dos elementos nuevos

   $R(AA,B,CC) = 0.75 \times 0.8 \times 0.9775 = 0.5865$

   $R(A,BB,CC) < 0.5$

   $R(A,BBB,C) < 0.5$

   $R(A,B,CCC) < 0.5$

4. Añadiendo tres elementos.

   $R(AAAA,B,C) = (1-(1-.5)^4) \times 0.8 \times 0.85 = .9375 \times 0.8 \times 0.85 = 0.6375$

   $R(AAA,BB,C)$=$0.875 \times 0.96 \times 0.85 = 0.714$ Este es el mejor

   $R(AAA,B,CC) = 0.875 \times 0.8 \times 0.85 = 0.7$

   $R(AA,BB,CC) = 0.75 \times 0.96 \times 0.9775 = 0.703$

   $R(AA,BBB,C) = 0.75 \times (1-(1-.8)^3) \times 0.85 = 0.6324$

   $R(AA,B,CCC) = 0.75 \times 0.8 \times (1-(1-.85)^3) = 0.597975$

   Los restantes no son mayores de 0.5 que sería la fiabilidad del primer componente, A, que no se repite.

Ejercicio 128 *El tiempo de vida de unos ciertos dispositivos sigue una distribución Normal de media 10000 horas y desviación típica 1000 horas.*

1. *Calcular la función de fiabilidad y la probabilidad de que uno de estos dispositivos dure al menos 9000 horas.*

2. *Si se sabe que uno de estos dispositivos ya ha durado 9000 horas, ¿Cuál es la probabilidad de que dure por lo menos 500 horas más?*

3. *Formamos un sistema en serie con dos dispositivos usados. El primero ha sido usado 9000 horas y el segundo 11000 horas. ¿ Cuál es la probabilidad de que este sistema dure 500 horas?*

4. *¿ Cuál es la probabilidad de que este sistema dure 500 horas si los dos dispositivos anteriores los colocamos en paralelo?*

1. $R(t) = \int_t^\infty \frac{1}{1000\sqrt{2\pi}} e^{-\frac{(t-10000)^2}{2\times 1000^2}} dt,$

   $R(9000) = \int_{9000}^\infty \frac{1}{1000\sqrt{2\pi}} e^{-\frac{(t-10000)^2}{2\times 1000^2}} dt = 0.841\,345$

   $\frac{9000-10000}{1000} = -1, \; 1 - F(-1) = 0.841\,345$

2. $\frac{P(x>9500)}{P(x>9000)} = \frac{0.691\,462}{0.841345} = .821\,853$

   $\frac{9500-10000}{1000} = -.0.5, \; 1 - F(-0.5) = .691\,462$

3. $\frac{11000-10000}{1000} = 1, \; 1 - F(1) = .158\,655$

   $\frac{11500-10000}{1000} = 1.5, \; 1 - F(1.5) = 6.680\,7 \times 10^{-2}$

   $\frac{P(x>11500)}{P(x>11000)} = \frac{6.680\,7\times 10^{-2}}{.158\,655} = .421\,08$

   El sistema en serie tiene de fiabilidad :

   $0.821\,853 \times 0.421\,08 = 0.346\,07$

4. $1 - (1 - 0.821\,853) \times (1 - 0.421\,08) = 0.896\,87$

Ejercicio 129 *Calcula:*

1. *La fiabilidad en un instante de un sistema como el siguiente si cada componente tiene en ese instante una fiabilidad de 0.4:*

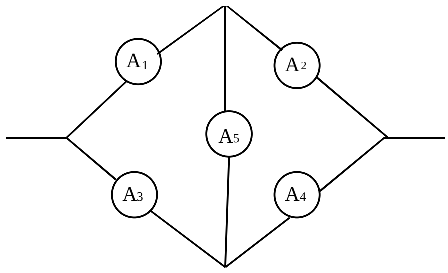


Figura 11.1

2. *Fiabilidad de un sistema formado por dos elementos en serie cada uno de ellos como el de la figura*

3. *Idem si se montan en paralelo*

1. $p_5 \times [(1-q_1q_3)(1-q_2q_4)] + q_5 \times [1-(1-p_1p_2)(1-p_3p_4)] =$
$0.4 \times (1-0.6^2)^2 + 0.6 \times \left[1-(1-0.4^2)^2\right] = 0.\,163\,84 + 0.\,176\,64 = 0.\,340\,48$

2. $0\,340\,48^2 = 0.\,115\,93$

3. $R_s(t) = 1-(1-0.\,340\,48)^2 = 0.\,565\,03.$

Ejercicio 130 *Hallar la fiabilidad de un flash con 3 pilas en redundancia secuencial con distribución de fallo exponencial.*

1. *Si se supone que las tres pilas son idénticos y la conmutación de un dispositivo a otro es perfecta:*

2. *Si se supone que las tasas de fallo son diferentes y la conmutación de un dispositivo a otro es perfecta:*

1. La fiabilidad se calcula considerando los tres casos posibles:

   a) Que el primer elemento sobreviva hasta t.

   b) Que falle el primero antes de t, pero que el segundo sobreviva desde ese momento hasta t.

   c) Que falle el primero y despues el segundo antes de t, pero que el tercero sobreviva desde el segundo fallo hasta por lo menos t.

   $e^{-\lambda t} + \int_0^t \lambda e^{-\lambda x} e^{-\lambda(t-x)} dx + \int_0^t \int_x^t \lambda e^{-\lambda x} \lambda e^{-\lambda(y-x)} e^{-\lambda(t-y)} dydx = e^{-t\lambda} + \lambda e^{-t\lambda} t + \frac{1}{2} t^2 \lambda^2 e^{-t\lambda}$

Otro procedimiento alternativo es considerar la variable aleatoria suma de los tres tiempos de duración de cada componente. Si esos tiempos son x, y, z, El tiempo de duración de sistema es $s = x+y+z$. Hallamos la función de infiabilidad correspondiente a esta variable:

$P(s \leq t) = \int_0^t \int_0^{t-x} \int_{x+y}^t \lambda^3 e^{-\lambda s} dsdydx = \int_0^t \int_0^{t-x} \left[-\lambda^2 e^{-\lambda s}\Big|_{x+y}^t \right] dydx =$

$= -\lambda^2 \int_0^t \int_0^{t-x} \left[\left(e^{-\lambda t}\right) - \left(e^{-\lambda x - \lambda y}\right) \right] dydx =$

$= -\lambda^2 \int_0^t \left[\left(e^{-\lambda t} y\right) + \frac{1}{\lambda} \left(e^{-\lambda x - \lambda y}\right)\Big|_0^{t-x} \right] dx =$

$= -\lambda^2 \int_0^t \left[\left(e^{-\lambda t}(t-x)\right) + \frac{1}{\lambda} \left(e^{-\lambda x - \lambda(t-x)}\right) - \frac{1}{\lambda} e^{-\lambda x} \right] dx =$

$= -\lambda^2 \int_0^t \left[\left(e^{-\lambda t} t - e^{-\lambda t} x)\right) + \frac{1}{\lambda} e^{-\lambda t} - \frac{1}{\lambda} e^{-\lambda x} \right] dx =$

$= -\lambda^2 \left(e^{-\lambda t} tx - e^{-\lambda t} \frac{x^2}{2} + \frac{1}{\lambda} e^{-\lambda t} x + \frac{1}{\lambda^2} e^{-\lambda x} \right)\Big|_{x=0}^{x=t} =$

$= \left(-\lambda^2 e^{-\lambda t} t^2 + \lambda^2 e^{-\lambda t} \frac{t^2}{2} - \lambda e^{-\lambda t} t - e^{-\lambda t} \right) + 1 =$

$= -e^{-t\lambda} - t\lambda e^{-t\lambda} - \frac{1}{2} t^2 \lambda^2 e^{-t\lambda} + 1$

Por tanto la función de Fiabilidad del sistema es:

$R_s(t) = 1 - P(s \leq t) =$

$= 1 - \left(-e^{-t\lambda} - t\lambda e^{-t\lambda} - \frac{1}{2} t^2 \lambda^2 e^{-t\lambda} + 1 \right) =$

$= e^{-t\lambda} + t\lambda e^{-t\lambda} + \frac{1}{2} t^2 \lambda^2 e^{-t\lambda}$

2. Si las tres pilas tienen distinta tasa de fallo

   $e^{-\lambda_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 x} e^{-\lambda_2 (t-x)} dx + \int_0^t \int_x^t \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 (y-x)} e^{-\lambda_3 (t-y)} dydx =$

   $e^{-\lambda_1 t} - \lambda_1 \frac{e^{-\lambda_1 t} - e^{-\lambda_2 t}}{\lambda_1 - \lambda_2} +$

   $+ \lambda_1 \lambda_2 \frac{e^{-\lambda_1 t} (\lambda_2 - \lambda_3) - e^{-\lambda_2 t} (\lambda_1 - \lambda_3) + e^{-\lambda_3 t} (\lambda_1 - \lambda_2)}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)}$

Ejercicio 131 *Un sistema esta compuesto por dos componentes en serie con tiempo de vida exponencial con una vida media de 200 horas y 500 horas respectivamente.*

1. *Hallar la función de fiabilidad del sistema.*
2. *Si ponemos ambos componentes en paralelo ¿Cual sería la función de fiabilidad del sistema?*

3. *Si añadimos en paralelo al sistema del apartado anterior un componente que sigue una distribución uniforme entre 0 y 150 horas ¿Aumenta la fiabilidad del sistema?*

4. *Si ponemos los tres elementos en paralelo ¿Cuál es la probabilidad de que este sistema de tres elementos en paralelo dure menos de 100 horas?*

1. La fiabilidad es el producto de las fiabilidades
   $e^{-(\frac{1}{200}+\frac{1}{500})t} = e^{-\frac{7}{1000}t}$

2. La infiabilidad es el producto de las infiabilidades, así que
   $R(t) = 1-(1-e^{-\frac{1}{200}t})(1-e^{-\frac{1}{500}t})$

3. Aumenta la fiabilidad.

4. Para el tercer componente la infiabilidad es
   $F(t) = \int_0^{100} \frac{1}{150} dt = .666\,6$
   Han de fallar los tres antes de las 100 horas
   La probabilidad de que fallen los tres $= (1-e^{-\frac{1}{200}100})(1-e^{-\frac{1}{500}100}).666\,67 = 4.\,754\,9\times 10^{-2}$

Ejercicio 132 *El tiempo de duración de ciertos componentes siguen una distribución exponencial con tasa de fallo de 0.005 fallos por hora.*
Se pide:

1. *Hallar la función de fiabilidad*

2. *Probabilidad que el componente dure menos de 300 horas.*

3. *Hallar la probabilidad de que dos de estos componentes no fallen antes de las 300 horas*

4. *Hallar la función de fiabilidad de tres de estos componentes colocados en paralelo*

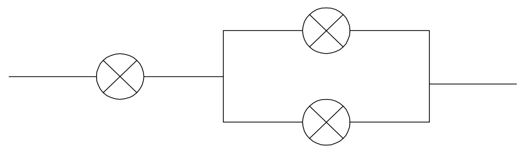
1. $R(t) = e^{-0.005t}$

2. $1-e^{-0.005t} = 1-e^{-0.005\times 300} = 0.\,776\,87$

3. $(1-0.\,776\,87)\,(1-0.\,776\,87) = 4.\,978\,7\times 10^{-2}$

4. $1-\left(1-e^{-0.005t}\right)\left(1-e^{-0.005t}\right)\left(1-e^{-0.005t}\right)$

Ejercicio 133 *Tres componentes identicos con fiabilidad exponencial y tiempo medio de vida de 2000 horas estan conectados formando el sistema de la figura*



1. *Hallar la función de fiabilidad de cada uno de sus componentes*
2. *Hallar la probabilidad de que cada componente dure al menos 1000 horas*
3. *Hallar la función de fiabilidad del sistema*
4. *¿Cúal es la probabilidad de que el sistema dure al menos 1000 horas.*

1. La función de fiabilidad exponencial es $R(t) = e^{-kt}$ siendo k la tasa de fallo que es la inversa de la vida media.

   Por lo tanto la función de fiabilidad de uno de estos componentes es $R(t) = e^{-\frac{1}{2000}t}$
2. $R(1000) = e^{-\frac{1}{2000}1000} = .606\,531$
3. La función de fiabilidad para un sistema en paralelo es $R_{par}(t) = 1 - \prod_{i=1}^{n}(1 - R_i(t))$

   En este caso será $1 - (1 - e^{-\frac{1}{2000}t})^2 = 2e^{-\frac{1}{2000}t} - e^{-\frac{1}{1000}t}$

   la función de fiabilidad de sistemas en serie es el producto de las fiabilidades. Por lo tanto el sistema total tendra la función de fiabilidad: $R_{sist}(t) = e^{-\frac{1}{2000}t}\left(2e^{-\frac{1}{2000}t} - e^{-\frac{1}{1000}t}\right) = 2e^{-\frac{1}{1000}t} - e^{-\frac{1}{1000}t}e^{-\frac{1}{2000}t} = 2e^{-\frac{1}{1000}t} - e^{-\frac{3}{2000}t}$
4. $R_{sist}(1000) = 2e^{-\frac{1}{1000}1000} - e^{-\frac{3}{2000}1000} = .512\,629$

Ejercicio 134 *Se supone que el vuelo de un avión es un sistema que consta de tres componentes principales: A (avión), B (tripulación) y C (aeropuerto), además el componente B puede considerarse como un subsistema en paralelo formado por un capitan (B_1) y un suboficial (B_2). También el aeropuerto consta de dos pistas (C_1,C_2) y el avión debe usar por lo menos una de ellas. Para que el vuelo se realice tienen que estar diponibles los tres componentes principales. La probabilidad de que cada cada uno de los elementos del sistema realice su función satisfactoriamente es la siguiente:*

$P(A) = 0.9999, P(B_1) = 0.995, P(B_2) = 0.8, P(C_1) = 0.95, P(C_2) = 0.85$

1. *¿Cuál es la fiabilidad del sistema?*
2. *¿Cuál sería si se añadiera una nueva pista de aterrizaje con probabilidad de estar utilizable el 50% de las veces?*
3. *¿Y si se suprimiera el suboficial?*

1. Sistema serie- paralelo. Hay tres componentes en serie. El segundo y el tercero son sistemas con dos componentes en paralelo.

   $R_S = R_A\,[1-(1-R_{B_1})(1-R_{B_1})]\,[1-(1-R_{C_1})(1-R_{C_2})] =$

   $0.9999\,[1-(1-0.995)(1-0.8)]\,[1-(1-0.95)(1-0.85)] = 0.991\,41$
2. $0.9999\times[1-(1-0.995)(1-0.8)]\times[1-(1-0.95)(1-0.85)(1-0.5)] = 0.995\,15$
3. $0.9999\,(1-(1-0.995))\,(1-(1-0.95)(1-0.85)) = 0.987\,44$

Ejercicio 135 *Cuatro unidades identicas permanecen en un sistema en redundancia activa con fallos independientes. Al menos tres de las unidades deben permanecer activas para que el sistema pueda cumplir su misión.*

1. *Si las unidades tienen función de fiabilidad exponencial con tasa de fallo 0.02, calcular la función de fiabilidad del sistema y su vida media.*
2. *¿Cual seria la la función de fiabilidad si solo se precisará una unidad para el funcionamiento del sistema? ¿Cuál sería en este caso la vida media del sistema?*

1. $R_s = \binom{4}{3}\left(e^{-0.02t}\right)^3(1-e^{-0.02t}) + \binom{4}{4}\left(e^{-0.02t}\right)^4 = 4.0\exp(-.0\,6t) - 3.0\exp(-.0\,8t)$

   La vida media puede obtenerse integrando la función de fiabilidad:

   $\int_0^\infty (4.0\exp(-.0\,6t) - 3.0\exp(-.0\,8t))\,dt = 29.\,167$
2. Sería un sistema en paralelo $1-(1-e^{-0.02t})^4$ .la vida media es en este caso

   $\mu_s = \mu + \frac{\mu}{2} + \frac{\mu}{3} + \frac{\mu}{4} = 50 + \frac{50}{2} + \frac{50}{3} + \frac{50}{4} = 104.\,167$

Ejercicio 136 *Hallar la función de fiabilidad, de distribución, de densidad y la vida media del sistema formado por dos componentes idénticos*

1. *Si ambos son exponenciales, están colocados en paralelo y cada uno de ellos tiene una vida media de 2 horas.*

2. *Si estan en paralelo y cada uno de ellos se rige por una distribución uniforme en el intervalo entre 0 y 4 horas.*

1. $f(t) = 1/2e^{-1/2\,t}, \;\; F(t) = 1 - e^{-1/2\,t}, \quad R(t) = e^{-1/2\,t}$

   $R_s(t) = 1 - \Pi_{i=1}^{n}[1 - R_i(t)] = 1 - \left(1 - e^{-\frac{1}{2}\,t}\right)^2 = 2e^{-\frac{1}{2}t} - e^{-t}$

   $F_s(t) = \left(1 - e^{-1/2\,t}\right)^2$

   $f(t) = \frac{d}{dt}\left(1 - e^{-\frac{1}{2}\,t}\right)^2 = -\left(-1 + e^{-\frac{1}{2}t}\right)e^{-\frac{1}{2}t} = e^{-\frac{1}{2}t} - e^{-t}$

   $\mu_s = 2 + 1 = 3\; horas$

2. $f(t) = 1/4, \;\; F(t) = 1/4\,t, \quad R(t) = 1 - 1/4\,t$

   $R_s(t) = 1 - \Pi_{i=1}^{n}[1 - R_i(t)] = 1 - [1 - (1 - 1/4\,t)]^2 = 1 - t^2/16$

   $F_s(t) = t^2/16,$

   $f_s(t) = t/8$

   $\mu_s = \int_0^4 t.\; t/8\; dt = \int_0^4 t^2/8\; dt = t^3/24\big|_0^4 = 64/24 = 2.66\; horas$

Ejercicio 137 *Un sistema esta compuesto por dos componentes en serie con tiempo de vida exponencial con una vida media de 100 horas y 400 horas respectivamente.*

1. *Hallar la función de fiabilidad del sistema y su vida media*

2. *Si ponemos ambos componentes en paralelo ¿Cual sería la función de fiabilidad del sistema?*

3. *Si añadimos al sistema del apartado anterior un componente en serie que sigue una distribución uniforme cuya vida está entre 0 y 150 horas ¿Aumenta la fiabilidad del sistema?*

4. *Si ponemos los tres elementos en paralelo ¿Cuál es la probabilidad de que el sistema dure menos de 100 horas?*

1. $\mu = \frac{1}{\lambda}, \; \lambda_1 = \frac{1}{100}, \; \lambda_2 = \frac{1}{400}$

   $R_s(t) = R_1(t)R_2(t) = e^{-\frac{1}{100}t}e^{-\frac{1}{400}t} = e^{-\frac{1}{80}t}$

   La vida media es por tanto 80 horas.

2. $R_s(t) = 1 - (1 - R_1(t))(1 - R_2(t)) = 1 - \left(1 - e^{-\frac{1}{100}t}\right)\left(1 - e^{-\frac{1}{400}t}\right) = e^{-\frac{1}{400}t} + e^{-\frac{1}{100}t} - e^{-\frac{1}{80}t}$

3. Como la fiabilidad de un sistema en serie es el producto de las fiabilidades y la fiabilidad de cada uno esta comprendido entre 0 y 1, el resultado no aumentaría la fiabilidad. Normalmente la disminuiría

4. La probabilidad de que el anterior componente (de distribución uniforme) dure menos de 100 horas es:
   $F_3(100) = \int_0^{100} \frac{1}{150} dt = \frac{100}{150} = \frac{2}{3}$
   La probabilidad de que el sistema anterior (de dos componentes exponenciales) dure menos de 100 horas es
   $F_s(100) = 1 - R_s(100) = 1 - \left(e^{-\frac{1}{400}100} + e^{-\frac{1}{100}100} - e^{-\frac{1}{80}100}\right) = 0.13982$
   Para que el sistema dure menos de 100 horas hace falta que los dos componentes duren menos de 100 horas, así que la probabilidad pedida será el producto de las probabilidades
   $0.13982 \times \frac{2}{3} = 9.3213 \times 10^{-2}$.

Ejercicio 138 *Un sistema consta de dos componentes idénticos con función de densidad exponencial y conectados en paralelo. La tasa de fallo en horas de cada componente es* 9×10^{-4}.

1. *Calcular la función de fiabilidad de cada componente.*
2. *Calcular la función de fiabilidad de un sistema con dos componentes en paralelo.*
3. *Calcular la probabilidad de que este sistema de dos componentes en paralelo dure al menos 1200 horas*
4. *¿Cúantos componentes como mínimo habría que colocar en paralelo para que la vida media del sistema sea al menos de 2400 horas?*

1. $R(t) = e^{-9\times10^{-4}t}$
2. $R(t) = 1 - \left(1 - e^{-9\times10^{-4}t}\right)^2$
3. $R(1200) = 1 - \left(1 - e^{-9\times10^{-4}\times1200}\right)^2 = 0.56387$

4. $\mu = \frac{1}{9\times10^{-4}.} = 1111.1$
$1111.1 + \frac{1111.1}{2} + \frac{1111.1}{3} + \frac{1111.1}{4} = 2314.8$
$1111.1 + \frac{1111.1}{2} + \frac{1111.1}{3} + \frac{1111.1}{4} + \frac{1111.1}{5} = 2537.0$
Como mínimo 5 componentes.

Ejercicio 139 *Tres componentes con tiempo de fallo exponencial y tasa de fallo 0.02, 0.04 y 0.05 (tiempo en horas) se han dispuesto formando un sistema en serie.*

1. *Cual la tasa de fallo del sistema.*
2. *Hallar la función de fiabilidad del sistema.*
3. *Calcular la probabilidad de que el sistema permanezca funcionando al menos 10 horas.*

1. Es la suma de las tasas de fallo de sus componentes: $0.02+0.04+0.05 = 0.11$
2. $R_s(t) = e^{-(0.02+0.04+0.05)t} = \exp -0.11t$
3. $R_s(10) == \exp(-0.11 \times 10) = 0.33287$.

Ejercicio 140 *Cinco unidades identicas permanecen en un sistema en redundancia activa con fallos independientes. Al menos dos de las unidades deben permanecer activas para que el sistema pueda cumplir su misión.*

1. *Si las unidades tienen función de fiabilidad exponencial con tasa de fallo 0.02, calcular la función de fiabilidad del sistema.*
2. *¿Cual seria la la función de fiabilidad si solo se precisará una unidad para el funcionamiento del sistema? ¿Y su vida media?*
3. *Si las cinco unidades estuvieran en redundancia secuencial y el dispositivo de conmutación fuese perfecto ¿Cual sería la vida media del sistema?*

1. $\sum_{i=k}^{n} \binom{n}{i} R(t)^i F(t)^{n-i} = \sum_{i=2}^{5} \binom{5}{i} e^{-0.02ti}(1-e^{-0.02t})^{5-i}$
$1 - F(t)^n = 1 - (1-e^{-0.02t})^5$
2. Es un sistema en paralelo. La vida media de cada componente es:
$$\frac{1}{0.02} = 50$$
La vida media del sistema es:
$$50 + 25 + \frac{50}{3} + 12.5 + 10 = 114.17$$

3. $50 \times 5 = 250.0$

Ejercicio 141 *Calcular la función de fiabilidad de dos componentes en redundancia secuencial con dispositivo conmutador perfecto si la duración de cada una de ellas se rige por una Distribución Uniforme en el intervalo [0,4]*

Las funciones de densidad, infiabilidad y fiabilidad de la duración de cada componente son:

$$f(t) = 1/4, \quad F(t) = 1/4\, t, \quad R(t) = 1 - 1/4\, t$$

Para calcular la función de fiabilidad del sistema distinguimos dos casos según el valor de t :

Realizamos este problema por dos métodos:

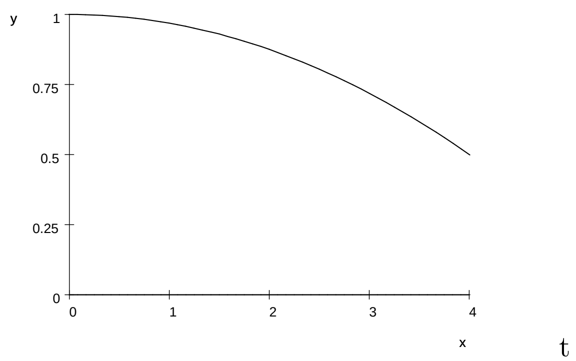
Primer método:

Primer caso, $t \leq 4$:

El primer componente puede durar hasta t, o el primero se rompe en para un tiempo x comprendido entre 0 y t y el segundo continúa funcionando desde x hasta t.

$R_s(t) = 1 - \frac{t}{4} + \int_0^t \frac{1}{4}\left[1 - \frac{1}{4}(t-x)\right] dx = 1 - \frac{t}{4} + \int_0^t \left[\frac{1}{4} - \frac{1}{16}t + \frac{1}{16}x\right] dx$ $= 1 - \frac{t}{4} + \int_0^t \left(\frac{1}{4}\right) dx - \int_0^t \left(\frac{1}{16}t\right) dx + \int_0^t \left(\frac{1}{16}x\right) dx = 1 - \frac{t}{4} + \frac{1}{4}t - \frac{1}{16}t^2 + \frac{1}{32}t^2 =$ $1 - \frac{1}{32}t^2$

Por tanto si $t \leq 4$ $\quad R_s(t) = 1 - \frac{1}{32}t^2$ cuya representación gráfica es la que sigue.



Segundo caso, $4 < t \leq 8$:

El primer componente no puede durar hasta t. Para que entre los dos cubran en funcionamiento un tiempo t,se han de cumplir los siguientes sucesos:

a) El primero tiene que durar al menos hasta $t - 4$.

b) El primero debe fallar para un x contenido entre $t-4$ y t, si ha durado hasta $t-4$.

c) El segundo debe durar al menos desde x hasta t.

La probabilidad de la intersección de estos sucesos resulta

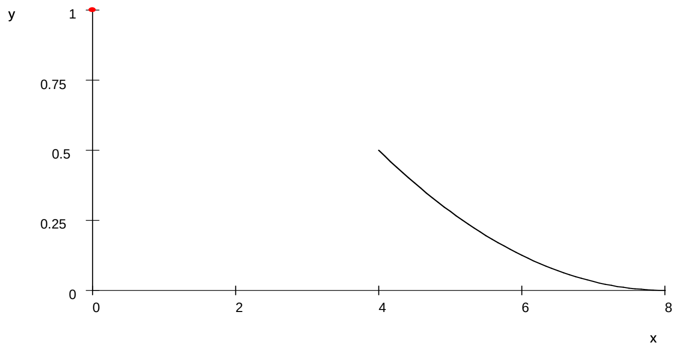
$R_s(t) = \left[1 - \frac{1}{4}(t-4)\right] \int_{t-4}^{t} \frac{\frac{1}{4}}{\frac{1}{4}[4-(t-4)]} \left[1 - \frac{1}{4}(t-x)\right] dx =$

$= \left[1 - \frac{1}{4}(t-4)\right] \frac{1}{8-t} \int_{t-4}^{4} \left[1 - \frac{1}{4}t + \frac{1}{4}x\right] dx =$

$= \frac{1}{4}\left(\int_{t-4}^{4} 1dx - \int_{t-4}^{4} \frac{1}{4}tdx + \int_{t-4}^{4} \frac{1}{4}xdx\right) =$

$= \frac{1}{4}\left(\frac{1}{8}t^2 - 2t + 8\right) = \frac{1}{32}t^2 - \frac{1}{2}t + 2$

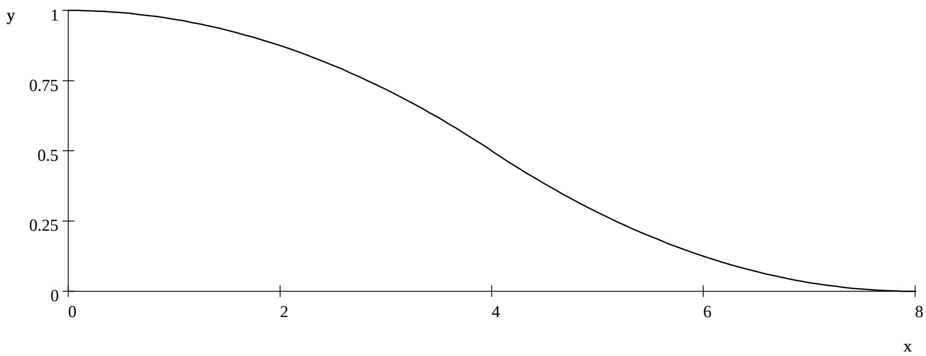
La representación gráfica es la que sigue.



La representación completa de la función de fiabilidad

$$R_s(t) = \begin{cases} 1 - \frac{1}{32}t^2 & \text{si} \quad 0 \le t \le 4 \\ \frac{1}{32}t^2 - \frac{1}{2}t + 2 & \text{si} \quad 4 \le t \le 8 \end{cases}$$

es:



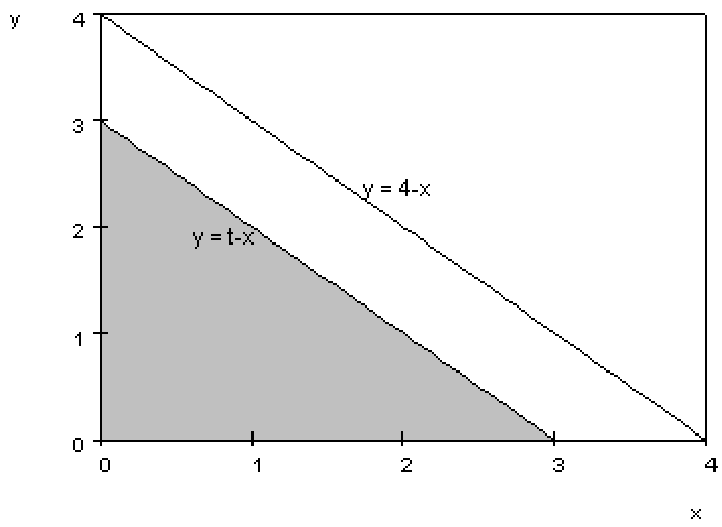
Segundo método:

Se considera x el tiempo que dura la primera componente, e y el tiempo que dura la segunda. Hallamos la función de infiabilidad de la variable $t = x + y$.

Para ello partimos de la función de densidad conjunta $f(x, y) = f(x) \times f(y) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$, considerando como independientes los tiempos hasta el fallo de ambas componentes.

Primer caso, $t \leq 4$:

Se calculará la función de infiabilidad por medio de integral de la función densidad en el intervalo comprendido entre $(0, t)$:



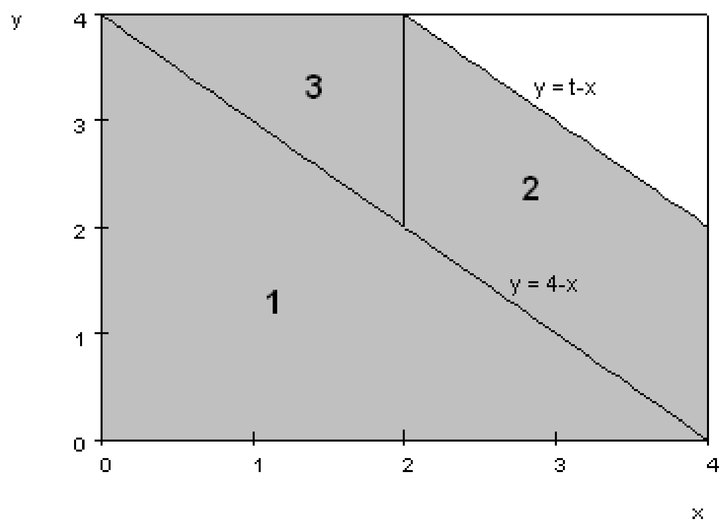
que corresponde a la región en gris de la figura:

$F(t) = \int\limits_{x=0}^{t} \int\limits_{y=0}^{t-x} \frac{1}{16}\ dy\ dx = \frac{1}{16} \int\limits_{x=0}^{t} \left[y|_0^{t-x} \right]\ dy\ dx = \frac{1}{16} \int\limits_{x=0}^{t} [t - x]\quad dx = \frac{1}{16} \int\limits_{x=0}^{t} \frac{[t-x]^2}{-2}\ dx = \frac{1}{32} t^2$

Por tanto la fiabilidad correspondiente a los valores de $t < 4$ sería:

$$R_s(t) = 1 - \frac{1}{32} t^2$$

Segundo caso, $4 < t \leq 8$:



En este caso la integral la hacemos en tres regiones.

La correspondiente a la región **1** es claramente 0.5. Puede hallarse tomando $t = 4$ en la expresión del primer caso:

$$F_1(4) = \tfrac{1}{32}4^2 = \tfrac{1}{2}$$

La correspondiente a la región **2** se obtiene con la integral:

$$F_2(t) = \int_{x=t-4}^{4} \int_{y=4-x}^{t-x} \tfrac{1}{16}\ dy\ dx = \int_{x=t-4}^{4} \left[\tfrac{1}{16} y\Big|_{4-x}^{t-x} \right]\ dx =$$

$$= \int_{x=t-4}^{4} \left[\tfrac{1}{16} \left((t-x) - (4-x)\right) \right]\ dx =$$

$$\tfrac{1}{16} \int_{x=t-4}^{4} \left((t-x) - (4-x)\right)\ dx = \tfrac{1}{16} \int_{x=t-4}^{4} (t-4)\ dx =$$

$$= \tfrac{t-4}{16} x\Big|_{t-4}^{4} = \tfrac{t-4}{16}(8-t)$$

La correspondiente a la región **3** se obtiene con la integral:

$$F_3(t) = \int_{x=0}^{t-4} \int_{y=4-x}^{4} \tfrac{1}{16}\ dy\ dx = \tfrac{1}{16} \int_{x=0}^{t-4} \left[y\Big|_{4-x}^{4} \right]\ dx =$$

$$= \frac{1}{16}\int\limits_{x=0}^{t-4} x\,dx = \frac{1}{16}\frac{x^2}{2}\Big|_0^{t-4} = \frac{1}{16}\frac{(t-4)^2}{2}$$

La función de infiabilidad sería la suma de las tres integrales:
$F_s(t) = \frac{1}{2} + \frac{t-4}{16}(8-t) + \frac{1}{16}\frac{(t-4)^2}{2} = \frac{1}{2}t - \frac{1}{32}t^2 - 1.$
Por tanto la función de fiabilidad para $4 < t \leq 8$ resulta:
$R_{s(t)} = 1 - \left(\frac{1}{2}t - \frac{1}{32}t^2 - 1\right) = \frac{1}{32}t^2 - \frac{1}{2}t + 2.$

Tercer método:

Adoptamos un enfoque geométrico, interpretando la función de distribución de probabilidad como volumen de un prisma cuya base es la región que cumple $x + y \leq t,$: y la altura el valor de la función densidad, $f(t) = \frac{1}{16}$.

Primer caso, $t \leq 4$:

La base es un triangulo de base t y de altura t. Asi que la probabilidad es $F_s(t) = \frac{t^2}{2} \times \frac{1}{16} = \frac{1}{32}t^2$. Por tanto en esta región $R_s(t) = 1 - \frac{1}{32}t^2$

Segundo caso, $4 < t \leq 8$:

La base puede calcularse restando del cuadrado completo el area de triángulo con $y > t - x$: $4^2 - \frac{(8-t)^2}{2} = 8t - \frac{1}{2}t^2 - 16$

Por tanto la probabilidad es $F_s(t) = \left(8t - \frac{1}{2}t^2 - 16\right) \times \frac{1}{16}. = \frac{1}{2}t - \frac{1}{32}t^2 - 1.$ En esta segunda región la fiabilidad es $R_{s(t)} = 1 - \left(\frac{1}{2}t - \frac{1}{32}t^2 - 1\right) = \frac{1}{32}t^2 - \frac{1}{2}t + 2.$

Ejercicio 142 *La vida media de un componente sigue una distribución exponencial de media 0.2 meses. Cuando este componente falla se reemplaza inmediatamente por otro idéntico, por lo que tiene que haber suficientes elementos de repuesto, ya que el suministrador sólo atiende la demanda una vez al mes. ¿Cuántos elementos hay que tener en stock si no se desea que el riesgo de quedarnos sin repuestos supere el 5%?*

La Distribución del número del número de elementos que fallen en un mes es una Poisson de parámetro $\frac{1}{0.2} = 5$. En este caso la probabilidad de que el número de fallos sea mayor que n es

$P(r > n) = \sum_{r=n+1}^{\infty} \frac{5^r}{r!}e^{-5} = 1 - \sum_{r=0}^{n} \frac{5^r}{r!}e^{-5}$

Si tenemos n componentes en stock, la probabilidad de que fallen más de n debe ser menor que 0.05:

$1 - \sum_{r=0}^{n} \frac{5^r}{r!}e^{-5} \leq 0.05; \quad \sum_{r=0}^{n} \frac{5^r}{r!}e^{-5} \geq 0.95$

Como $\sum_{r=0}^{8} \frac{5^r}{r!}e^{-5} = 0.931\,91$ y $\sum_{r=0}^{9} \frac{5^r}{r!}e^{-5} = 0.968\,17$, se deduce que el número de componentes almacenados debe ser al menos 9.

Este problema también puede resolverse usando la distribución de Erlang. De esta forma se ha hecho el ejercicio 143, que es similar a éste.

Ejercicio 143 *Trabajamos con un componente exponencial de vida media 0.2 meses. Cuando se rompe este componente debe sustituirse otro para poder seguir trabajando. Ocurre que la periodicidad del reparto de ese componente es mensual y por tanto solo podemos adquirir repuestos nuevos una vez al mes. Por ese motivo queremos tener en stock al menos el número de un número de elementos suficientes para ir reponiendo de modo que la probabilidad de quedarnos sin repuestos, y por tanto tener que detener la producción en medio del mes sea menor que 0,01. ¿De cuántos de estos componentes debemos disponer al principio de este ciclo mensual?*

Sea n el número de elementos de que dispongo al principio de mes. El tiempo hasta el fallo de estos n elementos (usados consecutivamentes, es decir en redundancia secuencial) es la suma de los tiempos hasta el fallo de cada uno de ellos independientemente. Por tanto sigue una distribución de Erlang. La función de fiabilidad de Erlang es:

$$R(t) = \int_t^\infty \frac{\lambda}{(n-1)!} (\lambda t)^{n-1} e^{-\lambda t} dt = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!}.$$

Como $\mu = \frac{1}{\lambda} = 0.2$, resulta para $\lambda = 5$. Como el conjunto de los n elementos debe durar al menos un mes tomamos $t = 1$.

Por lo tanto debemos determinar el menor valor de n que cumpla:

$$e^{-5} \sum_{i=0}^{n-1} \frac{(5)^i}{i!} > 1 - 0.01 = 0.99$$

Como para $n = 11$ resulta:

$e^{-5}(1+5+25/2+125/6+625/24+5^5/5!+5^6/6!+5^7/7!+5^8/8!+5^9/9!+5^{10}/(10)!) = 0.986\,3$

y para $n = 12$:

$$e^{-5}(1 + 5 + 25/2 + 125/6 + 625/24 + 5^5/5! + 5^6/6! + 5^7/7! + 5^8/8! + 5^9/9! + 5^{10}/(10)! + 5^{11}/(11)!) = .994\,55$$

resulta que debemos tener en total al principio del mes de al menos 12 elementos, incluido el que esté instalado al principio.

Si realizamos el ejercicio con la distribución de Poisson observamos que $\sum_{r=0}^{10} \frac{5^r}{r!} e^{-5} = 0.986\,3$, $\sum_{r=0}^{11} \frac{5^r}{r!} e^{-5} = 0.994\,55$ que tambien corresponden a 12 elementos, uno de ellos el que esté instalado al principio.

Ejercicio 144 *Cuando hay dos componentes en paralelo, parece razonable suponer que si uno de ellos falla el segundo esta sometido a unas condiciones más duras de trabajo, y por tanto tendrá más posibilidades de fallar:*

Supongamos dos componentes ídenticos con función de fiabilidad exponencial y colocados en paralelo. La tasa de fallo (en fallos cada mil horas) de cada elemento funcionando juntos es $\lambda_1 = 5$ y si sólo funciona uno de ellos es $\lambda_2 = 7$.

1. *Hallar la función de fiabilidad de un sistema de estas características.*

2. *Comparar la fiabilidad en el instante* $t = 100$ *horas de este sistema y de otro sistema en paralelo en forma convencional, en que la tasa de fallo del elemento que sobrevive continúe siendo 5.*

1. La probabilidad de que este sistema funcione al menos hasta un tiempo t es la suma de dos probababilidades, ya que el sistema funciona al menos hasta t si funcionan ambos elementos, o si habiendo fallado uno cualquiera de ellos antes de t el otro sobrevive desde el momento del fallo hasta t.

   La probabilidad de que funcionen ambos elementos hasta t es el producto de sus fiabilidades. $e^{-5t} \times e^{-5t}$. La probabilidad de que el primero de ellos falle en x, el segundo no falle en el intervalo de tiempo $(0, x)$ y continúe fucionando desde x hasta t es:

$$\int_0^t 5e^{-5x} e^{-5x} \times e^{-7(t-x)} dx$$

   La probabilidad de que el segundo de ellos falle en x, el primero no falle antes de x y continue la función del sistema hasta t toma este mismo valor, puesto que ambos componentes son iguales. Por tanto

$$R(t) = e^{-5t} \times e^{-5t} + 2 \int_0^t 5e^{-5x} e^{-5x} \times e^{-7(t-x)} dx = e^{-10t} + 10e^{-7t} \int_0^t e^{-3x} dx =$$

$$= e^{-10t} + 10e^{-7t} \left(\tfrac{1}{3} - \tfrac{1}{3}e^{-3t}\right) = e^{-10t} + \tfrac{10}{3}e^{-7t} - \tfrac{10}{3}e^{-10t} = \tfrac{10}{3}e^{-7t} - \tfrac{7}{3}e^{-10t}$$

2. En este sistema la fiabilidad en el instante $t = 100$ horas=0.1 miles de horas es:

   $R(0.1) = \frac{10}{3}e^{-0.7} - \frac{7}{3}e^{-1} = 0.796\,90$

   El sistema en paralelo habitual tiene de función de fiabilidad

$$R(t) = 1 - (1 - e^{-5t})^2$$

   Por tanto

   $R(0.1) = 1 - (1 - e^{-0.5})^2 = 0.845\,18$

   Como podía esperarse la fiabilidad es algo más alta en este caso, aunque seguramente más alejada de la realidad.

Ejercicio 145 *Se disponen de 7 elementos idénticos dispuestos en redundancia secuencial. La vida media de estos elementos es 1000 horas. Calcular, si no hay ningún problemas de conmutación:*

1. *La Fiabilidad del sistema para 3000 dias de funcionamiento y para 5000 dias.*

2. *La vida media del sistema de 7 elementos.*

1. La distribución del tiempo (t) transcurrido hasta el n-simo fallo sigue una distribución de Erlang. La fiabilidad puede calcularse por medio de la expresión:
$$R_s(t) = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!}$$
En este caso:
$R_s(3000) = e^{-0.001\times 3000} \sum_{i=0}^{7-1} \frac{(-0.001\times 3000)^i}{i!} = e^{-3} \sum_{i=0}^{6} \frac{(3)^i}{i!} = 0.966\,49$
$R_s(5000) = e^{-0.001\times 5000} \sum_{i=0}^{7-1} \frac{(-0.001\times 5000)^i}{i!} = e^{-5} \sum_{i=0}^{6} \frac{(5)^i}{i!} = 0.762\,18$

2. La vida media del sistema es, en este caso, la media de la distribución de Erlang:

$$n\mu = \frac{n}{\lambda} = n\mu = 7 \times 1000 = 7000 \ horas \tag{11.1}$$

T. 12

Inferencia con Pruebas de Vida

Ejercicio 146 *Calcular un estimador de máxima verosimilitud para el parámetro σ de una distribución normal, suponiendo que se conozca el parámetro μ. Emplear muestras con 3 elementos.*

El logaritmo neperiano de la función de verosimilitud es:

$$3\ln\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_3-\mu}{\sigma}\right)^2 =$$
$$3\ln\frac{1}{\sigma} + 3\ln\frac{1}{\sqrt{2\pi}} - \frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_3-\mu}{\sigma}\right)^2 =$$
$$3\ln\frac{1}{\sigma} + 3\ln\frac{1}{\sqrt{2\pi}} - \frac{1}{2\sigma^2}\left[(x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2\right]$$

La derivada de esta función con respecto a σ es:
$3\sigma\left(-\frac{1}{\sigma^2}\right) + 0 - \frac{1}{2}(-2)\sigma^{-3}\left[(x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2\right]$
Igualando esta derivada a 0 y multiplicando por σ^2, obtenemos la ecuación:

$$3\sigma + \sigma^{-1}\left[(x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2\right]$$

despejando σ, se halla su estimador de máximo verosimilitud, que resulta:

$$\widehat{\sigma} = \sqrt{\frac{(x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2}{3}}$$

Ejercicio 147 *Supongamos que se observan muestras de 50 elementos hasta que se obtenga el octavo fallo.*

Los tiempos de fallo han sido: 91, 145, 221, 285, 315, 328, 411, 496.

Estima el valor de la vida media de estos elementos bajo la hipótesis de distribución exponencial

1. Si la prueba de vida es con reposición

2. Si la prueba de vida es sin reposición

3. Hallar en cada caso un intervalo de confianza bilateral y unilateral al nivel 80%

1.
$$\widehat{\mu} = \frac{50 \times 496}{8} = \frac{24800.0}{8} = 3100.0$$

2.
$$\begin{aligned}\widehat{\mu} &= \frac{91 + 145 + 221 + 285 + 315 + 328 + 411 + 43 \times 496}{8} = \\ &= \frac{23124}{8} = 2890.5\end{aligned}$$

3. EL intervalo de confianza bilateral es para el primer caso:

$$\left(\frac{2T_{ac}}{\chi_{16,0.9}}, \frac{2T_{ac}}{\chi_{16,0.1}}\right) = \left(\frac{2 \times 24800.0}{23.5}, \frac{2 \times 24800}{9.31}\right) = (\ 2110.6\ ,\quad 5327.6\ )$$

El intervalo de confianza unilateral es

$$\left(\frac{2T_{ac}}{\chi_{16,0.8}}, \infty\right) = \left(\frac{2 \times 24800.0}{20.77}, \infty\right) = (2388.1, \infty)$$

Como el valor 80 no aparece en la tabla que se ha empleado, realizamos una iteración lineal entre 0.75 y 0.90, cuyos valores si vienen en la tabla, y son 19.4 y 23.5 respectivamente.

El valor obtenido es 20.77

Para el segundo caso, los intervalos correspondientes son:

$$\left(\frac{2T_{ac}}{\chi_{16,0.9}}, \frac{2T_{ac}}{\chi_{16,0.1}}\right) = \left(\frac{2 \times 23124}{23.5}, \frac{2 \times 23124}{9.31}\right) = (\ 1968.0,\quad 4967.6\ )$$

y

$$\left(\frac{2T_{ac}}{\chi_{16,0.8}}, \infty\right) = \left(\frac{2 \times 23124}{20.77}, \infty\right) = (\ 2226.7,\quad \infty\ )$$

Ejercicio 148 *Un fabricante nos informa que sus productos duran por término medio 10000 horas. Hemos instalado 50 unidades de este producto en nuestra empresa y al cabo de 990 horas hemos apreciado que habíamos tenido que reponer siete de ellas. ¿Podemos aceptar la información del fabricante al 95% de confianza?*

El valor estimado a partir de nuestra experiencia para la vida media es:

$$\widehat{\mu} = \frac{T_{ac}}{r} = \frac{50 \times 990}{7} = \frac{49500.0}{7} = 7071.4$$

Calculamos ahora un intervalo bilateral de confianza al 95% para la vida media
$\left(\frac{2\times 49500}{\chi_{14,0.975}}, \frac{2\times 49500}{\chi_{14,0.025}}\right) = \left(\frac{99000.0.}{26.119}, \frac{99000.0}{5.6287}\right) = (3790.3, \quad 17588.)$
Como este intervalo contiene el valor dado por el fabricante, damos por aceptable la información dada por éste.

Ejercicio 149 *Se someten 50 unidades a un ensayo censurado por número de fallos sin reposición. El ensayo se terminó al producirse el decimo fallo. Los tiempos hasta el fallo de los 10 elementos observados fueron:*

65, 110, 380, 420 ,505, 580, 650, 840, 910, 950.

Hallar un estimador para la vida media y un intervalo bilateral de confianza al 95% para este parámetro.

$65 + 110 + 380 + 420 + 505 + 580 + 650 + 840 + 910 + 950 + 40 \times 950 = 43410.0$

$\hat{\mu} = \frac{T_{ac}}{r} = \frac{\sum_1^r t_i + (n-r)t_r}{r} = \frac{65+110+380+420+505+580+650+840+910+950+40\times 950}{10} =$
$= 4341.0$

El intervalo de Confianza es
$\left(\frac{2\times 43410.0}{\left(\chi^2_{20}\right)^{-1}(0.975)}, \frac{2\times 43410.0}{\left(\chi^2_{20}\right)^{-1}(0.025)}\right) = (2540.9, 9052.4)$

Ejercicio 150 *Se someten 20 unidades a una prueba de vida hasta 10 fallos con reemplazamiento. El decimo fallo se ha producido a las 80 horas. Estimar la vida media , dando un intervalo de confianza bilateral al 95%*

$\hat{\mu} = \frac{T_{ac}}{r} = \frac{20\times 80}{10} = 160.0$
$\left(\frac{2\times 1600}{\left(\chi^2_{20}\right)^{-1}(0.975)}, \frac{2\times 1600}{\left(\chi^2_{20}\right)^{-1}(0.025)}\right) = \left(\frac{3200}{34.17}, \frac{3200}{9.5908}\right) = (93.649, 333.65)$

Ejercicio 151 *Se someten 20 unidades a una prueba de vida sin reemplazamiento durante un tiempo de 600 horas. En este intervalo de tiempo han fallado 18 de ellas. La duración de las unidades falladas (en horas) son:*

0.69, 0.94, 1.12, 6.79, 9.28, 9.31, 9.95, 12.9, 12.93, 21.33, 64.56, 69.66, 108.38, 124.88, 157.02, 190.19, 250.55, 552.87

Estimar la vida media, dando un intervalo de confianza bilateral al 95%.

$\hat{\mu} = \frac{T_{ac}}{K} = \frac{\sum_1^r t_i+(n-k)T}{K} = \frac{0.69+0.94+1.12+...+.19+250.55+552.87+2\times600}{18} = \frac{2803.4}{18} =$
$= 155.\,74$ horas.

Un intervalo bilateral (aproximado) de confianza para la vida media es.

$\left(\frac{2T_{ac}}{\chi_{38,0.975}}, \frac{2T_{ac}}{\chi_{36,0.025}}\right) = \left(\frac{2\times2803.\,4}{\left(\chi^2_{38}\right)^{-1}(0.975)}, \frac{2\times2803.\,4}{\left(\chi^2_{36}\right)^{-1}(0.025)}\right) = (98.55, 262.80).$

Ejercicio 152 *Se someten 2000 unidades a una prueba de vida sin reemplazamiento durante un tiempo de 600 horas. En este intervalo de tiempo han fallado 18 de ellas. La duración de las unidades falladas (en horas) son:*

0.69, 0.94, 1.12, 6.79, 9.28, 9.31, 9.95, 12.9, 12.93, 21.33, 64.56, 69.66, 108.38, 124.88, 157.02, 190.19, 250.55, 552.87

Estimar la vida media, dando un intervalo de confianza bilateral al 95%.

$\hat{\mu} = \frac{T_{ac}}{K} = \frac{\sum_1^r t_i+(n-k)T}{K} = \frac{0.69+0.94+1.12+...+.19+250.55+552.87+1982\times600}{18} = \frac{11908034}{18} =$
$= 6615\,60$ horas

Un intervalo bilateral (aproximado) de confianza para la vida media es.

$\left(\frac{2T_{ac}}{\chi_{38,0.975}}, \frac{2T_{ac}}{\chi_{36,0.025}}\right) = \left(\frac{2\times11908034}{\text{ChiSquareInv(0.975;38)}}, \frac{2\times11908034}{\text{ChiSquareInv(0.025;36)}}\right) =$
$= (418\ 590,\ \ 1\ 116\ 200).$

Ejercicio 153 *Se someten 20 unidades a una prueba de vida con reemplazamiento durante un tiempo de 600 horas. En este intervalo de tiempo han fallado 18 de ellas. La duración de las unidades falladas (en horas) son:*

0.69, 0.94, 1.12, 6.79, 9.28, 9.31, 9.95, 12.9, 12.93, 21.33, 64.56, 69.66,108.38, 124.88, 157.02, 190.19, 250.55, 552.87

Estimar la vida media, dando un intervalo de confianza bilateral al 95%.

$\hat{\mu} = \frac{T_{ac}}{K} = \frac{nT}{K} = \frac{20\times600}{18} = 666.\,67$ horas.

Un intervalo bilateral (aproximado) de confianza para la vida media es.

$\left(\frac{2T_{ac}}{\chi_{38,0.975}}, \frac{2T_{ac}}{\chi_{36,0.025}}\right) = \left(\frac{2\times12000}{\left(\chi^2_{38}\right)^{-1}(0.975)(0.975;38)}, \frac{2\times12000}{\left(\chi^2_{36}\right)^{-1}(0.025)}\right) =$
$= (421.\,83, 1124.\,9)$

Ejercicio 154 *Se ha realizado una prueba de vida con reemplazamiento de acuerdo con un plan de muestreo consistente en poner en funcionamiento 2500 elementos hasta que se produzca el 5º fallo. Se ha acordado que si la estimación de la vida media obtenida con esta prueba fuese mayor que 1500, se aceptaría el lote. Si el tiempo medio de vida ofrecido por el productor es de 3000 horas y el valor mínimo exigible por el comprador es de 1000 horas, calcular:*

1. *El riesgo del comprador y del productor.*

2. *El valor de la ordenada de la curva característica que corresponde al valor de 2500 horas para la vida media del producto.*

1. Riesgo del productor:
   $P(\hat{\mu} < C/\mu = \mu_0) = \alpha = P\left(\chi^2_{2r} < \frac{2rC}{\mu_0}\right) =$
   $\alpha = P\left(\chi^2_{10} < \frac{2\times5\times1500}{3000}\right) = P\left(\chi^2_{10} < 5.0\right) = 0.108$
   Riesgo del comprador:
   $\beta = 1 - P\left(\chi^2_{2r} < \frac{2rC}{\mu_0}\right) = 1 - P\left(\chi^2_{10} < 15\right) = 1 - 0.87 = 0.13.$

2. La ordenada de la curva caractéristica correspondiente al valor 2500 para la vida media es la probabilidad de aceptar un lote cuya vida media fuera 2500 horas.
   $P_{AC}(\mu = 2500) = P\left(\chi^2_{2r} > \frac{2\times5\times1500}{2500}\right) = 1 - P\left(\chi^2_{10} < 6.0\right) = 1 - 0.18 =$ 0.82

Ejercicio 155 *Se desea estimar la duración de un cierto tipo de lámparas. Para ello se ha observado la duración de 15 de ellas, hasta que han fallado todas. Se supone que los tiempos de vida se ajustan bien a una distribución normal. Los tiempos de vida de estas lamparas, en horas, resultaron ser*
848, 932, 938, 959, 961, 993, 1120, 1126,
1012, 1013, 1035, 1066, 1085, 1123, 1166.
Calcular:

1. *Estimación de la vida media e intervalos unilateral y bilateral de confianza al 95%.*

2. *Estimación de la desviación típica y un intervalo de confianza para ésta al 90%.*

1. $\widehat{\mu} = \frac{849+932+....+1123+1166}{15} = 1025.1$
   $s = \sqrt{\frac{(849-1025.1)^2+(932-1025.1)^2+....+(1123-1025.1)^2+(1166-1025.1)^2}{14}} = 89.214$
   El intervalo bilateral de confianza sigue la expresión:
   $$\left(\widehat{\mu} - t_{n-1,1-\frac{\alpha}{2}} s/\sqrt{n} \le \mu \le \widehat{\mu} + t_{n-1,1-\frac{\alpha}{2}} s/\sqrt{n}\right)$$
   $\left(1025.1 - t_{14,0.975} 89.214/\sqrt{15} \le \mu \le 1025.1 + t_{14,0.975} 89.214/\sqrt{15}\right);$
   $(1025.1 - 2.1448 \times 23.035 \le \mu \le 1025.1 + 2.1448 \times 23.035);$

$(1025.1 - 49.405 \leq \mu \leq 1025.1 + 49.405)$; $975.70 \leq \mu \leq 1074.5$

Para el intervalo unilateral de confianza tomamos el de extremo inferior que es el que se usa frecuentemente

$$\left(\widehat{\mu} - t_{14,0.95}\ s/\sqrt{n} \leq \mu\right)$$

$(1025.1 - 1.7613 \times 23.035 \leq \mu) = (984.53 \leq \mu)$

2. La estimación para la desviación típica es $s = 89.214$. Un intervalo bilateral de confianza para la desviación típica es:

$$s\sqrt{\frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}} \leq \sigma \leq s\sqrt{\frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}}}$$

$89.214\sqrt{\frac{14}{\chi^2_{14,0.975}}} \leq \sigma \leq 89.214\sqrt{\frac{14}{\chi^2_{14,0.025}}}$

$89.214\sqrt{\frac{14}{26.119}} \leq \sigma \leq 89.214\sqrt{\frac{14}{5.6287}}$

$65.316 \leq \sigma \leq 140.70.$

Ejercicio 156 *En nuestra empresa empleamos unos dispositivos electrónicos cuya función de densidad de probabilidad de su duración sin fallos (en horas) es* $f(t) = 7 \times 10^{-4} e^{-7\times 10^{-4} t}$, $t > 0$.

Por otra parte hemos hecho una test de vida con otros dispositivos del mismo tipo que nos ofrece un nuevo proveedor. Dicho test ha consistido en poner en funcionamiento 20 de estas unidades hasta que fallarán diez de ellos, en una prueba sin reposición. El registro del tiempo de fallos hasta que ha ocurrido el décimo fallo ha sido: 940, 950, 951, 970, 982, 1007, 1021, 1050, 1079, 1154. ¿Pueden considerarse estos nuevos dispositivos más fiables al 95% de confianza?

La vida media de los dispositivos que usamos en la actualidad es de 1428.57 horas (inversa de la tasa de fallo). Para los dispositivos nuevos hallamos, en primer lugar, la estimación de la vida media, que resulta:

$\widehat{\mu} = \frac{940+950+951+970+982+1007+1021+1050+1079+11\times 1154}{10} = 2164.4$

Esta estimación es mayor que la vida media de nuestros dispositivos. Sin embargo, realizando un intervalo de confianza unilateral:

$$\left(\frac{2\times 21644}{\left(\chi^2_{20}\right)^{-1}(0.95)}, \infty\right) = \left(\frac{2\times 21664}{31.41}, \infty\right) = (\ 1378.2,\ \ \infty\ )$$

encontramos que, al 95% de confianza, la vida media de los nuevos dispositivos será mayor que 1378.2 horas. Así que los nuevos dispositivos no pueden considerarse más fiables que los que usamos en la actualidad. ya que 1378<1428.57.

Ejercicio 157 *En nuestra empresa empleamos unos dispositivos electrónicos cuya función de densidad de probabilidad de su duración sin fallos (en horas) es* $f(t) = 7.5 \times 10^{-4} e^{-7.5\times 10^{-4} t}$, $t > 0$.

Por otra parte hemos hecho una test de vida con otros dispositivos del mismo tipo que nos ofrece un nuevo proveedor. Dicho test ha consistido en poner en funcionamiento durante 1200 horas 20 de estas unidades en una prueba sin reposición. El registro del tiempo de fallos hasta que ha transcurrido las 1200 horas fué: 940, 950, 951, 970, 982, 1007, 1021, 1050, 1079, 1154. ¿Pueden considerarse estos nuevos dispositivos más fiables al 95% de confianza?

La vida media de los dispositivos que usamos en la actualidad es de 1333.33 horas (inversa de la tasa de fallo). Para los dispositivos nuevos hallamos, en primer lugar, la estimación de la vida media, que resulta:
$\widehat{\mu} = \frac{20\times 1200}{10} = 2400$ horas

Realizando un intervalo unilateral de confianza al 95% para la vida media,basado en la prueba de fiabilidad realizada encontramos que es:
$\left(\frac{2\times 24000}{\left(\chi^2_{22}\right)^{-1}(0.95)}, \infty \right) = (1414.9, \infty)$

Como la vida media de nuestros dispositivos es 1333.33<1414.9, adoptaremos la decisión de que los nuevos dispositivos son más fiables.

Unidad Temática IV

PROBLEMAS DE ANÁLISIS DE LA VARIANZA

T. 13

Análisis de varianza con un factor

Ejercicio 158 [1] *Un vendedor de refrescos esta considerando la importancia del color del bote en la cantidad de ventas. El registro del número de unidades vendidas en diferentes tiendas de la ciudad elegidas al azar es el siguiente:*

Azul (X)	*93, 85, 89*
Rojo (Y)	*102, 86, 90, 100, 89, 94*
Amarillo (Z)	*81, 82, 80, 84*

1. *¿Se debe concluir que el color tiene alguna influencia sobre la cantidad promedio de unidades vendidas?*

2. *Hallar un intervalo de confianza (al 95%) para la diferencia de las medias de ventas entre los botes rojos y amarillos*

1. Calculamos las medias y las sumas de cuadrados en la siguiente tabla de valores:

[1] En todos los problemas se supondrá que se cumplen las hipótesis de partida válidas para aplicar el análisis de varianza.

Color	Valor	media color	Media total
X	93	89	88.846
X	85	89	88.846
X	89	89	88.846
Y	102	93.5	88.846
Y	86	93.5	88.846
Y	90	93.5	88.846
Y	100	93.5	88.846
Y	89	93.5	88.846
Y	94	93.5	88.846
Z	81	81.75	88.846
Z	82	81.75	88.846
Z	80	81.75	88.846
Z	84	81.75	88.846
$g.\ l.$			

SCF	SCR	STC
$(89 - 88.846)^2$	$(93 - 89)^2$	$(93 - 88.846)^2$
$(89 - 88.846)^2$	$(85 - 89)^2$	$(85 - 88.846)^2$
$(89 - 88.846)^2$	$(89 - 89)^2$	$(89 - 88.846)^2$
$(93.5 - 88.846)^2$	$(102 - 93.5)^2$	$(102 - 88.846)^2$
$(93.5 - 88.846)^2$	$(86 - 93.5)^2$	$(86 - 88.846)^2$
$(93.5 - 88.846)^2$	$(100 - 93.5)^2$	$(90 - 88.846)^2$
$(93.5 - 88.846)^2$	$(93.5 - 93.5)^2$	$(102 - 88.846)^2$
$(93.5 - 88.846)^2$	$(89 - 93.5)^2$	$(89 - 88.846)^2$
$(93.5 - 88.846)^2$	$(94 - 93.5)^2$	$(94 - 88.846)^2$
$(91.75 - 88.846)^2$	$(81 - 81.75)^2$	$(81 - 88.846)^2$
$(91.75 - 88.846)^2$	$(82 - 81.75)^2$	$(82 - 88.846)^2$
$(91.75 - 88.846)^2$	$(80 - 81.75)^2$	$(80 - 88.846)^2$
$(91.75 - 88.846)^2$	$(84 - 81.75)^2$	$(84 - 88.846)^2$
$g.l. = 2$	$g.l. = 10$	$g.l. = 12$
$\sum_{i=1}^{g} n_i \left(\overline{y}_i - \overline{y}\right)^2$	$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(y_{ij} - \overline{y}\right)^2$	$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(y_{ij} - \overline{y}_i\right)^2$
$= 331.442$	$= 244.425$	$= 575.692$

Tabla de Análisis de la varianza:

variación	suma de cuadrados	g. libertad
entre grupos	331.442	2
dentro de los grupos	244.25	10
Total	575.692	12

medias de cuadrados	$\mathrm{F}_{\exp}$	$\mathrm{F}_{2,12}^{-1}(0.95)$
$\frac{331.442}{2} = 165.72$	$\frac{165.72}{24.425} = 6.7849$	4.1028
$\frac{244.25}{10} = 24.425$	24.425	

Como $F_{\exp} > F_{2,12}^{-1}(0.95)$; $6.7849 > 4.1028$, se concluye que parece que el color del bote tiene influencia en las cantidades vendidas del refresco

2. Para calcular el intervalo de confianza para la diferencia entre las medias, al 95% sustituimos los datos de los botes rojos y amarillos en:

$$\left(|\overline{y}_p - \overline{y}_q| - t_{n-g}\sqrt{MCR\left(\frac{1}{n_p} + \frac{1}{n_q}\right)}, |\overline{y}_p - \overline{y}_q| + t_{n-g}\sqrt{MCR\left(\frac{1}{n_p} + \frac{1}{n_q}\right)}\right)$$

$t_{10}^{-1}(0.975) = 2.228$

$$\left(|93.5\text{-}81.75| - 2.228\sqrt{24.43\left(\frac{1}{6}+\frac{1}{4}\right)}, |93.5\text{-}81.75| + 2.228\sqrt{24.43\left(\frac{1}{6}+\frac{1}{4}\right)}\right) = (4.64,\ 18.86)$$

Como el intervalo de confianza para la diferencia de medias no contiene el valor 0, se concluye que las ventas con botes rojos parecen superar a las ventas con botes amarillos.

Ejercicio 159 *Los siguientes datos dan el consumo de electricidad diario por habitante realizado en 4 barrios de una ciudad. Los distintos datos provienen de 6 mediciones seleccionados al azar entre las realizadas en los días de un año .*

Barrio A	*Barrio B*	*Barrio C*	*Barrio D*
13.1	*11.4*	*10.6*	*11.5*
13.4	*12.1*	*11.1*	*12.0*
13.8	*12.1*	*11.4*	*12.9*
14.4	*12.6*	*12.5*	*13.4*
14.0	*12.8*	*11.7*	*12.6*
14.8	*13.4*	*13.0*	*14.0*
13.9167	*12.4*	*11.7167*	*12.7333*

1. *¿ Se puede considerar diferente el consumo medio por barrio.?*

2. *Calcular intervalos de confianza para la diferencia entre las medias de consumo entre los barrios usando el método de Bomferroni*

3. *Se observa que la media en el barrio A es superior a la del barrio C. Es esta diferencia significativa al nivel de significación 0.05?*

1. SE construye la tabla de Análisis de varianza:

Fuente de Variación	Suma de cuadrados	g.l.	medias cuadráticas	F_{exp}
Debida al Factor	15.2283	3	5.07611	$\frac{5.07611}{0.6245}$ $= 8.1283$
Residual	12.49	20	0.6245	
Total	27.7183	23		

$8.1283 > F^{-1}_{3,20}(0.95) = 3.0984$, por lo tanto no puede admitirse la igualdad de consumo entre los 4 barrios.

2. Hay que establecer $\binom{4}{2} = 6$ comparaciones. El nivel de significación de cada comparación es $\frac{\alpha}{6} = \frac{0.05}{6} = 8.3333 \times 10^{-3}$ por tanto el valor de t_{20} es.

   $t^{-1}_{20}(1 - 8.3333 \times 10^{-3}/2; 20) = t^{-1}_{20}(0.99583) = 2.9268$. El intervalo de confianza para la diferencia entre los barrios A Y B es

   $$\left(|\overline{y}_p - \overline{y}_q| - t_{n-g}\sqrt{MCR\left(\frac{1}{n_p} + \frac{1}{n_q}\right)}, |\overline{y}_p - \overline{y}_q| + t_{n-g}\sqrt{MCR\left(\frac{1}{n_p} + \frac{1}{n_q}\right)}\right) =$$

   $$\left(|13.917\text{-}12.4| - 2.927\sqrt{0.6245\left(\frac{1}{6} + \frac{1}{6}\right)}, |13.917\text{-}12.4| + 2.927\sqrt{0.625\left(\frac{1}{6} + \frac{1}{6}\right)}\right)$$
   $= (0.18134, 2.8521)$

   Los intervalos de confianza para la diferencia de consumo promedio entre las otras cinco comparaciones son los siguientes:

   Barrios A y C: $(2.2 - 1.33551, 2.2 + 1.33551) = (0.86449,\ 3.5355)$

   Barrios A y D : $(1.18333 - 1.33551,\ 1.18333 + 1.33551) = (-0.15218,$ $2.5188)$

   Barrios B y C : $(0.68333 - 1.33551, 0.68333 + 1.33551) = (-0.65218, 2.$ $0188\ )$

   Barrios B y D : $(-1.33551, -0.33333 + 1.33551) = (-1.3355, 1.0022)$

   Barrios C y D : (-1.01667-1.33551,-1.01667+1.33551) = (-2.3522, 0.31884)

3. El intervalo de confianza para la diferencia de los valores medios entre A y C no contiene el valor 0, así que no puede admitirse que los consumos sean similares, así que se decide que los consumos en los barrios A y c son signifícativamente diferentes.

Ejercicio 160 *El beneficio obtenido (en millones de pesetas) por cinco supermercados en distintos años viene dado en la siguiente tabla*

Super. 1	*Super. 2*	*Super. 3*	*Super. 4*	*Super. 5*
222	*196*	*204*	*305*	*128*
220	*235*	*190*	*351*	*109*
170	*188*	*182*	*351*	*112*
175		*190*	*348*	*139*
155		*104*		*70*

1. *Hacer la tabla de Análisis de Varianza.*
2. *¿Hay evidencia suficiente para concluir que el beneficio es distinto en algunos de estos supermercado?*
3. *Si es así, índica cuales son y por qué motivo.*

1.

Super. 1	Super. 2	Super. 3	Super. 4	Super. 5	
222	196	204	305	128	
220	235	190	351	109	
170	188	182	351	112	
175		190	348	139	
155		104		70	
188.4	206.33	174	338.75	111.6	197.45

$SCF = 5\,(188.4 - 197.45)^2 + 3\,(206.33 - 197.45)^2 + 5\,(174 - 197.45)^2 + 4\,(338.75 - 197.45)^2 + 5\,(111.6 - 197.45)^2 = 120111$

$STC = (222 - 197.55)^2 + (220 - 197.45)^2 + ... + (139 - 197.45)^2 + (70 - 197.45)^2 = 4165.8 + 1501.4 + 9125.5 + 81388 + 39608 =$

$1\,357\,95$

$SCR = 135795 - 120111 = 15684.$

Variación	Suma de cuadrados	g. l.	medias de cuadrados	F_{exp}
Factor	120111	4	30028	32.55
Residuales	15684	17	922.577	
Total	135795	21		

2. La $F_{4,17}$ teórica correspondiente a la significación 0.05 es 2.96.<<32.55. Por lo tanto se rechaza la hipótesis de igualdad de las medias entre los supermercados.

3. Usando intervalos de confianza individuales al 95%

$$\left(\overline{y_i} - t_{n-g,1-\frac{\alpha}{2}}\sqrt{\frac{\text{MCR}}{n_i}} < \mu_i < \overline{y_i} + t_{n-g,1-\frac{\alpha}{2}}\sqrt{\frac{\text{MCR}}{n_i}}\right)$$

Supermercado 1:
$\left(188.4 - \text{TInv}(0.975;17)\sqrt{\frac{922.58}{5}} < \mu_i < 188.4 + \text{TInv}\,(0.975;17)\sqrt{\frac{922.58}{5}}\right) =$
$\left(188.4 - 2.109\,8\sqrt{\frac{922.58}{5}}, 188.4 + 2.109\,8\sqrt{\frac{922.58}{5}}\right) = (159.\,74, 217.\,06)$
Supermercado 2:
$\left(206.33 - \text{TInv}(0.975;17)\sqrt{\frac{922.58}{3}} < \mu_i < 206.33 + \text{TInv}(0.975;17)\sqrt{\frac{922.58}{3}}\right) =$
$\left(206.33 - 2.109\,8\sqrt{\frac{922.58}{3}}, 206.33 + 2.109\,8\sqrt{\frac{922.58}{3}}\right) = (169.\,33, 243.\,33)$
Supermercado 3:
$\left(174 - \text{TInv}(0.975;17)\sqrt{\frac{922.58}{5}} < \mu_i < 174 + \text{TInv}(0.975;17)\sqrt{\frac{922.58}{5}}\right) =$
$\left(174 - 2.109\,8\sqrt{\frac{922.58}{5}}, 174 + 2.109\,8\sqrt{\frac{922.58}{5}}\right) = (145.\,34, 202.\,66)$
Supermercado 4:
$\left(338.75 - \text{TInv}(0.975;17)\sqrt{\frac{922.58}{4}} < \mu_i < 338.75 + \text{TInv}(0.975;17)\sqrt{\frac{922.58}{4}}\right) =$
$\left(338.75 - 2.109\,8\sqrt{\frac{922.58}{4}}, 338.75 + 2.109\,8\sqrt{\frac{922.58}{4}}\right) = (306.\,71, 370.\,79)$
Supermercado 5:
$\left(111.6 - \text{TInv}(0.975;17)\sqrt{\frac{922.58}{5}} < \mu_i < 111.6 + \text{TInv}(0.975;17)\sqrt{\frac{922.58}{5}}\right) =$
$\left(111.6 - 2.109\,8\sqrt{\frac{922.58}{5}}, 111.6 + 2.109\,8\sqrt{\frac{922.58}{5}}\right) = (82.\,941, 140.\,26)$

Puede observarse que los intervalos de confianza de los supermercados 4 y 5 no tienen parte en comun ni entre sí ni.con los restantes Por tanto concluimos que el 4 es el que obtiene más beneficio y el 5 el que obtiene menos. Los restantes intervalos de confiamza si tienen parte en común. Admitimos por tanto que los supermercados 1, 2 y 3 consiguen por término medio iguales beneficios.

Ejercicio 161 *Los datos siguientes se refieren a las pérdidas de peso de ciertas piezas mecánicas debidas a la fricción cuando la usaron tres fabricantes diferentes*

Fabricante A	*12.2, 11.8, 13.1,11.0, 3.9, 4.1, 10.3, 8.4*
Fabricante B	*10.9, 5.7, 13.5, 9.4, 11.4, 15.7, 10.8, 14.0*
Fabricante C	*12.7, 19.9, 13.6, 11.7, 18.3, 14.3, 22.8, 20.4*

Probar al nivel de significación 0.01 si las diferencias entre las medias de desgaste entre los fabricantes es signifícativa.

Las medias parciales son respectivamente 9.35, 11.245, 16.7125. La media total es 12.4658

$STC = (12.2-12.49)^2+(11.8-12.49)^2+(12.3.1-12.49)^2+....+(22.8-12.49)^2+(20.4-12.49)^2 = 507.46$

$SCF = 8(9.35-12.495)^2+8(11.425-12.495)^2+8(16.7125-12.49)^2 = 230.58$

$SCR = 507.46 - 230.58 = 276.88$

La tabla de análisis de la varianza es

Fuente de Variación	Sumas de cuad.	g.l.	medias cuadráticas	F_{exp}
Factor	230.58	2	115.29	8.7478
Residuos	276.88	21	13.18	
Total	507.46	23		

El valor de $F_{2,21}$ correspondiente al nivel de significación del test 0.01 es 5.78. El valor de F experimental supera al teórico, por lo que se rechaza la hipótesis nula. No se puede suponer que el desgaste sea el mismo en los tres fabricantes.

Ejercicio 162 *La siguiente tabla recoge el número de disquetes defectuosas fabricados usando diferentes sistemas de fabricación durante seis meses consecutivos.*

Sistema A	*Sistema B*	*Sistema C*
6	*14*	*10*
14	*9*	*12*
10	*12*	*7*
8	*10*	*15*
11	*14*	*11*
8	*12*	*11*

1. *Puede detectarse alguna diferencia significativa entre el número de defectuosos que produce cada sistema de fabricación*

2. *Hallar un intervalo de confianza para la media de defectos mensuales obtenidos con el Sistema A*

3. *Hallar un intervalo de confianza para la diferencia entre las media de defectos mensuales obtenidos con el Sistema A Y B.*

1. media total = 10.78

   Media Sistema A = 9.5

   Media Sistema B =: 11.833

Media Sistema C = 11.0

$STC = (6-10.78)^2 + (14-10.78)^2 + ... + (11-10.78)^2 + (11-10.78)^2 = 111.11$

$SCR = (6-9.5)^2 + ... + (8-9.5)^2 + (14-11.83)^2 + ... + (12-11.83)^2 + (10-11)^2 + +... + (11-11)^2 = 39.5 + 20.833 + 34.0 = 94.333$

$SCF = 111.111 - 94.333 = 16.778$

Tabla de análisis de varianza:

Fuente de Variación	Suma de cuadrados	g.l.	medias cuadráticas	F_{exp}
Debida al Factor	16.778	2	8.389	$\frac{8.389}{6.289} = 1.3349$
Residual	94.333	15	6.289	
Total	111.111	17		

El $P - value$ para esta F_{exp} es 0.2930, por lo tanto al nivel 0.05 no puede rechazarse la igualdad entre las medias. Se concluye que no hay diferencia significativa entre los tres sistemas.

2. Como hemos concluido que todos los sistemas son iguales podemos usar la muestra total para el intervalo de confianza, quedando en este caso.

$\left(10.778 - t_{17}^{-1}(0.975)\frac{2.5565}{\sqrt{18}}, 10.778 + t_{17}^{-1}(0.975)\frac{2.5565}{\sqrt{18}}\right) =$

$(9.5067, 12.049)$

$t_{17}^{-1}(0.975) = 2.1098$

Si de todas formas usamos el intervalo de confianza individual (Usando los datos del grupo y respetando la hipótesis de igualdad de la varianza) se obtendría

$\left(9.5 - t_{15}^{-1}(0.975)\frac{\sqrt{6.2888889}}{\sqrt{6}}, 9.5 + t_{15}^{-1}(0.975)\frac{\sqrt{6.2888889}}{\sqrt{6}}\right) =$

$(7.3178, 11.682) = (7.3178, 11.682)$.

$t_{15}^{-1}(0.975) = 2.1314$

3. Usando el estádistico

$t_{n-g} = \frac{|\overline{y}_p - \overline{y}_q|}{\sqrt{MCR\left(\frac{1}{n_p}+\frac{1}{n_q}\right)}}$, y sin usar la conclusión dada en 1, tenemos que el intervalo es

$$\left(|\overline{y}_p - \overline{y}_q| - t_{n-g}\sqrt{MCR\left(\frac{1}{n_p}+\frac{1}{n_q}\right)}, |\overline{y}_p - \overline{y}_q| + t_{n-g}\sqrt{MCR\left(\frac{1}{n_p}+\frac{1}{n_q}\right)}\right).$$

Al 95% de confianza sería

$$\left(|9.5-11.8333|-2.13\sqrt{6.29\left(\tfrac{2}{6}\right)},|9.5-11.8333|+2.13\sqrt{6.29\left(\tfrac{2}{6}\right)}\right)$$

$$=(-.752\,739,5.\,419\,34)$$

En el caso de que quisieramos considerar comparaciones multiples usando la desigualdad de Bonferroni, tomamos $\alpha^*=\frac{0.05}{3}=1.\,666\,67\times 10^{-2}$; $1-\frac{\alpha^*}{2}=1-\frac{1.\,666\,67\times10^{-2}}{2}=.\,991\,667$

$t_{15}^{-1}(0.991667)=2.\,693\,8$

El intervalo de confianza sería entonces

$$\left(|9.5\text{-}11.8333|-2.69\sqrt{6.29\left(\tfrac{1}{6}+\tfrac{1}{6}\right)},|9.5\text{-}11.8333|+2.69\sqrt{6.29\left(\tfrac{1}{6}+\tfrac{1}{6}\right)}\right)$$

$$=(-1.\,561\,8,6.\,228\,4)$$

Ejercicio 163 *Un experimento consiste en determinar el efecto de las burbujas de aire en la resistencia del asfalto. Las burbujas de aire se controlan en tres niveles: Bajo(2% - 4%), Medio(4% - 6%) y Alto(6%-8%).*

Los datos medidos sobre la resistencia del asfalto en los distintos niveles son:

Burbujas	*Resistencia del asfalto*
Bajo	*106 90 103 90 79 88 92 95*
Medio	*80 69 94 91 70 83 87 83*
Alto	*74 80 62 69 76 85 69 85*

¿Los niveles de burbujas de aire influyen en la resistencia del asfalto a un nivel de significación 0.01?

Fuente de Variación	Suma de cuadrados	g.l.	medias cuadráticas	F_{exp}
Debida al Factor	1295.58	2	647.792	$\frac{647.792}{73.7976}= \\ =8.\,78$
Residual	1549.75	21	73.7976	
Total	2845.3323	23		

Como la $F_{2,12}^{-1}(0.99)=5.\,780\,4$.verifica que $8.78>5.7804$, los niveles de burbujas de aire influyen en la resistencia del asfalto.

Ejercicio 164 *La siguiente tabla recoge el número de errores cometidos por cuatro cajeras de un supermercado en cinco meses consecutivos.*

Cajera A	*Cajera B*	*Cajera C*	*Cajera D*
6	*14*	*10*	*9*
14	*9*	*12*	*12*
10	*12*	*7*	*8*
8	*10*	*15*	*10*
11	*14*	*11*	*11*

1. *Hacer la tabla de análisis de la varianza ¿Puede atribuirse al azar la diferencia de errores entre estas cajeras?*
2. *Halla intervalo de confianza individuales para la media de errores cometidos por la primera cajera*
3. *Halla un intervalo de confianza para la diferencia promedio de errores cometidos por las cajeras A y B*
4. $6 + 14 + 10 + 8 + 11 = 49.0/5 = 9.8$

1. $14 + 9 + 12 + 10 + 14 = 59.0/5 = 11.8$

Cajera A	Cajera B	Cajera C	Cajera D	Total
9. 8	11. 8	11.0	10.0	10.65

$STC = 114.55$

$SCF = 5(9.8 - 10.65)^2 + 5(11.8 - 10.65)^2 + 5(11 - 10.65)^2 +$

$+5(10 - 10.65)^2 = 12.95$

$SCR = 101.6$

$F_{\text{exp}} = \frac{\frac{12.95}{3}}{\frac{101.6}{16}} = \frac{4.3167}{6.35} 0.67979$

$F^{-1}_{3,16}(0.95) = 3.2389.$

Como la F experimental es menor que la teórica se acepta la hipótesis de igualdad de medias, por lo que la diferencia entre ellas puede atribuirse al azar.

2. $\left(9.8 - \text{TInv}(0.975; 16)\sqrt{\frac{6.35}{5}}, 9.8 + \text{TInv}(0.975; 16)\sqrt{\frac{6.35}{5}}\right) =$

$\left(9.8 - 2.1199\sqrt{\frac{6.35}{5}}, 9.8 + 2.1199\sqrt{\frac{6.35}{5}}\right) = (7.4110, 12.189)$

3. $\left(2. - \text{TInv}(0.975;16)\sqrt{\frac{6.35\times2}{5}}, 2+\text{TInv}(0.975;16)\sqrt{\frac{6.35\times2}{5}}\right) =$
$\left(2. - 2.1199\sqrt{\frac{6.35\times2}{5}}, 2+2.1199\sqrt{\frac{6.35\times2}{5}}\right) = (-1.3786, 5.3786)$

Ejercicio 165 *Se estudia la valoración de los estudiantes de distintos lugares de procedencia sobre la calidad de la residencia universitaria donde habitan, sus valoraciones se recogen en la siguiente tabla:*

Procedencia	*Valoración*
Sevilla	*7,5,6,8*
Resto deEspaña	*6, 8,7,7*
Europa	*5,4,4,5*
América	*7,4,4,7*

1. *Hallar la tabla de análisis de varianza*

2. *¿La valoración media depende del lugar de origen?*

1.

Procedencia	Valoración	medias
Sevilla	7,5,6,8	6.5
Resto deEspaña	6, 8,7,7	7
Europa	5,4,4,5	4.5
América	7,4,4,7	5.5
Total		5.875

$$\begin{aligned} STC &= SCF + SCR \\ \sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2 &= \sum_{i=1}^{g} n_i(\overline{y}_i-\overline{y})^2 + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2 \end{aligned}$$

(a) $STC = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2 =$
$(7-5.875)^2+(5-5.875)^2+...+(4-5.875)^2+(7-5.875)^2 = 31.75$
$SCF = \sum_{i=1}^{g} n_i(\overline{y}_i-\overline{y})^2 =$
$4(6.5-5.875)^2+4(7-5.875)^2+4(4.5-5.875)^2+4(5.5-5.875)^2 = 14.75$
$SCR = STC - SCF = 31.75 - 14.75 = 17.0$
La tabla de análisis de la varianza resulta:

	Suma de cuadrados	g.l	medias	F_{exp}
Factor	14.75	3	$\frac{14.75}{3} = 4.917$	$\frac{4.9167}{1.4167} = 3.470$
Residuales	17	12	$\frac{17}{12} = 1.417$	
Total	31.75	15		

$\mathrm{FInv}(0.95; 3, 12) = 3.490\,3 > 3.470\,5.$

(a) Se acepta la hipótesis de igualdad entre las medias, por lo tanto se concluye que no hay influencia de la procedencia del estudiante en su valoración de la residencia en la que habita.

Ejercicio 166 *Para ver si el precio de un producto alimenticio depende del barrio en que se adquiere, se ha seleccionado al azar un número de tiendas de cada barrio y el precio de este producto en cada una de estas tiendas se ha registrado en la siguiente tabla*

Barrio A	*Barrio B*	*Barrio C*
210	*182*	*226*
192	*200*	*198*
183	*187*	*185*
227	*182*	*237*
242		*237*
212		

1. *Construir la tabla de análisis de la varianza*

2. *Contrastar la igualdad entre los precios medios del producto en los tres barrios*

1. $\overline{y_A} = \frac{210+192+183+227+242+212}{6} = 211.0$

   $\overline{y_B} = \frac{182+200+187+182}{4} = 187.\,75$

   $\overline{y_C} = \frac{226+198+185+237+237}{5} = 216.\,6$

   $\overline{y} = \frac{211\times 6+187.75\times 4+216.6\times 5}{15} = 206.\,67$

   $STC = (210-206.67)^2 + (192-206.67)^2 + (183-206.67)^2 + (227-206.67)^2 + (242-206.67)^2 + (212-206.67)^2 + (182-206.67)^2 + (200-206.67)^2 + (187-206.67)^2 + (182-206.67)^2 + (226-206.67)^2 + (198-206.67)^2 + (185-206.67)^2 + (237-206.67)^2 + (237-206.67)^2 = 6883.\,3$

   $SCF = 6\,(211-206.67)^2 + 4\,(187.75-206.67)^2 + 5\,(216.6-206.67)^2 = 2037.\,4$

   $SCR = 6883.\,3 - 2037.\,4 = 4845.\,9$

   La tabla de Análisis de la varianza es:

F.Variación	Suma de cuadrados	g.l	Media de cuadrados	$\mathrm{F}_{\exp}$
Factor	2037.4	2	$\frac{2037.4}{2} = 1018.7$	$\mathrm{F}_{2,12} = \frac{1018.7}{403.83} = 2.5226$
Residual	4845.9	12	$\frac{4845.9}{12} = 403.83$	
Total	6883.3	14		

2. Comparando la F experimental anterior con la F teórica $\mathrm{FDist}(0.95; 2, 12)$

   $F_{2,12}^{-1}(0.95) = 3.8853 > 2.5226$

   Por tanto, con el 95% de confianza, no se puede rechazar la hipótesis de igualdad de las medias en los tres barrios, así que concluimos que el precio del producto no depende del barrio.

T. 14

Análisis de varianza con varios factores

Ejercicio 167 [1]*En un experimento para investigar la calidad de un plástico se ha medido la resistencia del material en diferentes condiciones de temperatura y de humedad, Se han realizado dos medidas de la resistencia del material obtenido para cada combinación de temperatura y humedad. Los valores obtenidos se muestran en la tabla siguiente:*

	temperatura 1	*temperatura 2*
Humedad 1	*13, 11*	*8, 9*
Humedad 2	*11, 10*	*6, 7*

1. *Halla la tabla del análisis de varianza usando un modelo con interacción.*

2. *¿Es la interacción significativa?*

3. *¿Tiene influencia en la resistencia del plástico los cambios de temperatura?*

4. *¿Tienen influencia los cambios de humedad?*

Tabla de las medias:

	temperatura 1	temperatura 2	
Humedad 1	12 (2)	8.5 (2)	10.25 (4)
Humedad 2	10.5 (2)	6.5 (2)	8.5(4)
	11.25 (4)	7.5 (4)	Total= 9.375(8)

[1]En todos los problemas se supondrá que se cumplen las hipótesis de partida necesarias para poder aplicar el análisis de varianza.

Los valores entre paréntesis son el número de datos empleados en cada media.

Tenemos que calcular las smas de cuadrados:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}\left(y_{ijk}-\overline{y}\right)^2=\sum_{i=1}^{a}n_{i\cdot}\left(\overline{y}_{i\cdot}-\overline{y}\right)^2+\sum_{i=1}^{b}+\text{n}_{\cdot j}\left(\overline{y}_{\cdot j}-\overline{y}\right)^2$$
$$\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}\left(\overline{y}_{ij}-\overline{y}_{i\cdot}-\overline{y}_{\cdot j}+\overline{y}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}\left(y_{ijk}-\overline{y}_{ij}\right)^2$$

Los cálculos necesarios están organizados en las siguiente tabla:

		Fact. humedad	Fact. temp
y_{ijk}	$\left(y_{ijk}-\overline{y}\right)^2$	$\text{n}_{i\cdot}\left(\overline{y}_{i\cdot}-\overline{y}\right)^2$	$\text{n}_{\cdot j}\left(\overline{y}_{\cdot j}-\overline{y}\right)^2$
13	$(13-9.375)^2$	$(10.25-9.375)^2$	$(11.25-9.375)^2$
11	$(11-9.375)^2$	$(10.25-9.375)^2$	$(11.25-9.375)^2$
8	$(8-9.375)^2$	$(10.25-9.375)^2$	$(7.5-9.375)^2$
9	$(9-9.375)^2$	$(10.25-9.375)^2$	$(7.5-9.375)^2$
11	$(11-9.375)^2$	$(8.5-9.375)^2$	$(11.25-9.375)^2$
10	$(10-9.375)^2$	$(8.5-9.375)^2$	$(11.25-9.375)^2$
6	$(6-9.375)^2$	$(8.5-9.375)^2$	$(7.5-9.375)^2$
7	$(7-9.375)^2$	$(8.5-9.375)^2$	$(7.5-9.375)^2$
Total	37.875	6.125	28.125
g.l	n-1=7	a-1=1	b-1=1

Interacción	Residuos
$n_{ij}\left(\overline{y}_{ij}-\overline{y}_{i\cdot}-\overline{y}_{\cdot j}+\overline{y}\right)^2$	$\left(y_{ijk}-\overline{y}_{ij}\right)^2$
$(12-10.25-11.25+9.375)^2$	$(13-12)^2$
$(12-10.25-11.25+9.375)^2$	$(11-12)^2$
$(8.5-10.25-7.5+9.375)^2$	$(8-8.5)^2$
$(8.5-10.25-7.5+9.375)^2$	$(9-8.5)^2$
$(10.5-8.5-11.25+9.375)^2$	$(11-10.5)^2$
$(10.5-8.5-11.25+9.375)^2$	$(10-10.5)^2$
$(6.5-8.5-7.5+9.375)^2$	$(6-6.5)^2$
$(6.5-8.5-7.5+9.375)^2$	$(7-6.5)^2$
0.125	3.5
$(a-1)(b-1)=1$	$n-ab=4$

Tabla de análisis de la varianza:

F. de variación	S.cuad.	g.l	M.cuad.	F.exp.	F.teórica(0.05)
Humedad	6.125	1	6.125	$\frac{6.125}{0.875}=7.0$	7.7086
Temperatura	28.125	1	28.125	$\frac{28.125}{0.875}=32.143$	7.7086
Interacción	0.125	1	0.125	$\frac{0.125}{0.875}=0.14286$	7.7086
residuos	3.5	4	0.875		

Por lo tanto la Interacción no es signíficativa (no hay interacción). No influye la humedad, pero si influye la temperatura.

Si decidimos pasar al modelo sin interacción obtenemos el siguiente resultado

$Fuente\ de\ variación$	$S.cuad.$	$g.l$	$M.cuad.$
$Humedad$	6.125	1	6.125
$Temperatura$	28.125	1	28.125
$residuos$	3.625	5	$\frac{3.625}{5} = 0.725$

$F.exp.$	$F.teórica(0.05)$
$\frac{6.125}{.725} = 8.4483$	7.7086
$\frac{28.125}{.725} = 38.793$	7.7086

Con este modelo y para el mismo nivel de significación (0.05) concluiremos que tanto la temperatura como la humedad tienen influencia.

Ejercicio 168 *Se realiza un experimento para estudiar la influencia de los neumáticos de los coches en el desgaste de las pastillas de freno. Para ello se realizan recorridos con 5 marcas de neumáticos. Como se sospecha que el desgaste debe tener relación con el tipo de suelo, se realizan recorridos idénticos por suelo asfaltado de autopista, carretera comarcal y camino rural. El resultado de la prueba se resume en la siguiente tabla donde la variable respuesta es una medida del desgaste.*

	Comarcal	*Autopista*	*Rural*	*Medias por aditivo*
Neumático 1	*14.4*	*10.6*	*18.8*	*14.6*
Neumático 2	*11.3*	*5.5*	*9.9*	*8.9*
Neumático 3	*7.4*	*2.2*	*7.1*	*5.6*
Neumático 4	*10.7*	*5.5*	*10.6*	*8.9*
Neumático 5	*13.5*	*11.6*	*15.5*	*13.5*
Medias por modelos	*11.5*	*7.1*	*12.4*	*10.3*

1. *¿Influye la marca de los neumáticos en el desgaste? Utiliza el modelo de bloques completos al azar*

2. *¿Es aceptable utilizar el tipo de carretera como bloque?*

1. Consideramos el modelo de dos factores sin interacción:

$$STC = SCF_A + SCF_B + SCR$$

$$STC = (14.4 - 10.3)^2 + (11.3 - 10.3)^2 + ... + (13.5 - 10.3)^2 +$$

$+(10.6-10.3)^2++...+(11.6-10.3)^2+(18.8-10.3)^2+...+$

$+(15.5-10.3)^2=259.87$

$SCF_{Neum}=3\,(14.6-10.3)^2+3\,(8.9-10.3)^2+3\,(5.6-10.3)^2+$

$+3\,(8.9-10.3)^2+3\,(13.5-10.3)^2=164.\,22$

$SCF_{carret}=5(11.5-10.3)^2+5(7.1-10.3)^2+5(12.4-10.3)^2=80.45$

$SCR=259.87-(164.\,22+80.\,45)=15.\,2$

La tabla de análisis de la varianza es:

Fuente de Variación	Suma de cuad.	g.l.	medias. cuad.	F
Neumáticos	164.22	4	41.055	21.594
Carretera	80.45	2	40.22	21.154
Residuos	15.2	8	1.901	
Total	259.87	14		

Contraste para los neumáticos:

La F teórica para $\alpha=0.05$, $F_{4,8}^{-1}(0.95)=3.84<21.594$. Se rechaza la hipótesis de igualdad de las medias para los neumáticos, por lo que los neumáticos influyen en el desgaste.

Contraste para las carreteras:

La F teórica para $\alpha=0.05$, $F_{2,8}^{-1}(0.95)=4.46<21.154$,por lo tanto influye el tipo de carretera, así que es aceptable utilizar el tipo de carretera como bloque.

Ejercicio 169 *Para estudiar el consumo de aceite de un motor se prueban 4 motores distintos con 3 tipos de aceite obteniéndose 12 medidas de consumo. Se han obtenido los resultados siguientes :*

Suma de los cuadrados del factor aceite=100
Suma de los cuadrados del factor motor=80
Suma total de cuadrados=220
Se pide:

1. *Escribe la tabla ANOVA, considerando que no hay interacciones entre los factores.*

2. *¿Se puede considerar que los tipos de aceite no tienen influencia en el consumo?, ¿Y el tipo de motor? (Utiliza un nivel de significación 0.05).*

1.

FUENTE	Sum. de Cuad.	g.l..	Medias de Cuad.	F_{EXP}	F_{TEOR}
aceite	100	2	50	$\frac{50}{6.6667} = 7.5$	5.14
motor	80	3	26.667	$\frac{26.6667}{6.6667} = 4$	4.76
Residual	40	6	6.6667		
Total	220	11			

2. A la vista de los valores anteriores se concluye que los tres aceites no se consumen por igual ($7.5 > 5.14$), pero los motores probados pueden considerarse equivalentes con respecto al consumo de aceite ($4 < 4.76$).

Ejercicio 170 *El beneficio obtenido (en millones de pesetas) por cinco supermercados en cinco años viene dado en la siguiente tabla*

	Super. 1	*Super. 2*	*Super. 3*	*Super. 4*	*Super. 5*
1990	*222*	*196*	*204*	*305*	*128*
1991	*220*	*235*	*190*	*351*	*109*
1992	*170*	*188*	*182*	*351*	*112*
1993	*175*	*199*	*190*	*348*	*139*
1994	*155*	*108*	*104*	*205*	*70*

1. *Hacer la tabla de Análisis de Varianza usando el año como bloque.*

2. *¿Hay evidencia suficiente para concluir que hay algún supermercado con beneficio significativamente mayor que los demás? ¿Y menor?*

1.

FUENTE	Sum. de Cuad.	g.l..	Medias de Cuad..	F_{EXP}	F_{TEOR}
año	28135.	4	7033.84	12.04	0.0001
supermercado	106111.0	4	26527.8	45.40	0.0000
Residual	9349.84	16	584.365		
Total	143597	24			

2. El valor tan pequeño del $p-value$ correspondiente a los supermercados indica que hay diferencias significativas entre los beneficios obtenidos por estos. Para localizar cuales son los responsables de estas diferencias utilizo el Método de Bomferroni para hacer intervalos de confianza simultaneos para las medias de beneficios de cada uno de supermercado.

Como estimacion de la varianza del beneficio de todos los supermercados empleo la media de los cuadrados de los residuales. El nivel de significación para cada intervalo de confianza es $\frac{\alpha}{5} = 0.01$. Realizando los intervalos de confianza en orden creciente del beneficio medio, obtenemos:

(a) Intervalo de confianza para el supermercado 5:
$\left(111.6 - t_{16}^{-1}(0.995)\frac{\sqrt{584.365}}{\sqrt{5}}, 111.6 + t_{16}^{-1}(0.995)\frac{\sqrt{584.365}}{\sqrt{25}}\right) =$
$=(111.6 - 2.920\,8 \times 10.\,811, 111.6 + 2.920\,8 \times 10.\,81)$:
$=(111.6 - 31.\,576, 111.6 + 31.\,576) = (80.\,024, 143.\,18)$

(b) Intervalo de confianza para el supermercado 3:
$(174 - 31.\,576, 174 + 31.\,576) = (142.\,42, 205.\,58)$

(c) Intervalo de confianza para el supermercado 2:
$(185.2 - 31.\,576, 185.2 + 31.\,576) = (153.\,62, 216.\,78)$

(d) Intervalo de confianza para el supermercado 1:
$(188.4 - 31.\,576, 188.4 + 31.\,576) = (156.82, 219.98)$

(e) Intervalo de confianza para el supermercado 4:
$(312 - 31.\,576, 312 + 31.\,576) = (280.42, 343.58)$
Por lo tanto el supermercado 4 obtiene más beneficio que el resto, ya que su intervalo de confianza no tiene parte común con ninguno de los otros cuatro. En cambio no puede decirse que el supermercado cinco obtenga menos beneficio que todos los demás., ya que su intervalo de confianza, aunque por poco margen tiene parte común con el supermercado 3. Si obtiene menos beneficio que los supermercados 2, 1 y 4.

Ejercicio 171 *Los siguientes datos son los tiempos empleados por tres trabajadores usando tres tipos de maquinaría diferente en dos días*

	Trabajador 1	*Trabajador 2*	*Trabajador 3*
Maquinaria A	*37, 43*	*38, 44*	*38, 40*
Maquinaria B	*31, 36*	*40, 44*	*43, 41*
Maquinaria C	*36,40*	*33, 37*	*41, 39*

Usando un modelo de dos factores (maquinaria y trabajador) con interacción

1. *Hacer la tabla de Análisis de la varianza*

2. *¿Algun tipo de maquinaría es más rápida que las demás? ¿Algun trabajador es más rápido?*

3. ¿Hay interacción entre los factores?

1. Tabla de medias

	Trabajador 1	Trabajador 2	Trabajador 3	
Maquinaria A	40.0	41.0	39.0	40
Maquinaria B	33. 5	42.0	42.0	39.167
Maquinaria C	38.0	35.0	40.0	37.667
	37.18	39.333	40.333	38.943

$STF_A = 6 \times (40 - 38.943)^2 + 6 \times (39.167 - 38.943)^2 + 6 \times (37.667 - 38.943)^2 = 16.774$

$SCF_B = 6 \times (37.18 - 38.943)^2 + 6 \times (39.333 - 38.943)^2 + 6 \times (40.333 - 38.943)^2 = 31.154$

$SCR = (37-40)^2+(43-40)^2+(38-41)^2+(44-41)^2+(38-39)^2+(40-39)^2+(31-33.5)^2+(36-33.5)^2+(40-42)^2+(44-42)^2+(43-42)^2+(41-42)^2+(36-38)^2+(40-38)^2+(33-35)^2+(37-35)^2+(41-40)^2+(39-40)^2 = 78.5$

$STC = (37-38.943)^2+(43-38.943)^2+(38-38.943)^2+(44-38.943)^2+(38-38.943)^2+(40-38.943)^2+(31-38.943)^2+(36-38.943)^2+(40-38.943)^2+(44-38.943)^2+(43-38.943)^2+(41-38.943)^2+(36-38.943)^2+(40-38.943)^2+(33-38.943)^2+$
$(37-38.943)^2+(41-38.943)^2+(39-38.943)^2 = 220.94$

Tabla ANOVA

F.de Var.	S. de c.	g.l.	Media cuadrática	F_{exp}	F teórica
(A)	16.774	2	$\frac{16.774}{2} = 8.39$	$\frac{8.39}{8.72} = 0.962$	4. 256
(B)	31.154	2	$\frac{31.154}{2} = 15.58$	$\frac{15.58}{8.72} = 1.79$	4. 256 5
Inter(AB)	94.512	4	$\frac{94.512}{4} = 23.63$	$\frac{23.63}{8.72} = 2.71$	3. 633
Res(R)	78.5	9	$\frac{78.5}{9} = 8.72$		
Total(T)	220.94	17			

2. Todos los trabajadores y todas las maquinas son similares

3. No hay interacción entre los factores.

Ejercicio 172 *Se desea adquirir un nuevo equipo de maquinas para una fábrica. Para comparar las velocidades de la nueva maquinaria con la antigua. se encarga un cierto trabajo a 5 empleados que ya lo han realizado con el equipo antiguo. La tabla siguiente resume los tiempos en minutos.*

Empleado	*1*	*2*	*3*	*4*	*5*
Equipo nuevo	*115*	*205*	*147*	*121*	*186*
Equipo antiguo	*124*	*212*	*151*	*132*	*195*

¿A qué conclusión puede llegarse

1. *Usa el modelo de dos factores sin interacción (El factor bloque es el empleado) Tomese para alfa el valor 0.05*

2. *Usa el test de dos muestras con datos pareados*

1. Tabla de medias

Empleado	1	2	3	4	5	medias
Equipo nuevo	115	205	147	121	186	154.8
Equipo antiguo	124	212	151	132	195	162.8
medias	119.5	208.5	149	126.5	190.5	158.8

$STC = 12492.$

$SCF_{eq} = 5(154.8 - 158.8)^2 + 5\,(162.8 - 158.8)^2 = 160.0$

$SCF_{emp} = 2(119.5 - 158.8)^2 + 2\,(208.5 - 158.8)^2 + 2(149 - 158.8)^2 + 2\,(126.5 - 158.8)^2 + 2(190.5 - 158.8)^2$

$= 12318.$

$SCR = 12492. - 12318 - 160 = 14.0$

$F_{equipo} = \frac{\frac{160}{1}}{\frac{14}{4}} = 45.714$

El valor de la F teorica es $F_{F(1,4)}^{-1}(0.95) = 7.7086 < 45.714$, por lo que se rechaza la hipótesis nula de igualdad en los equipos, concluyendo por tanto que el equipo nuevo es más rápido. El *p-value* del test es $1 - F_{F(1,4)}(45.714) = 1 - 0.9975 = 0.0025$

2. Analizando la diferencia entre ambas muestras obtenemos la muestra $\{-9,\ -7,\ -4,\ -11,\ -9\}$. Se cálcula el estadístico del test :

$t = \frac{\overline{d}-0}{s/\sqrt{n}} = \frac{-8}{\sqrt{7/5}} = -6.7612$

El p-value del test es $2 * F_{t(4)}(-6.7612; 4)) = 2.5 \times 10^{-3} = 0.0025$.

La conclusión es la misma que en el caso anterior, ya que el *p-value* es el mismo. No es de extrañar, ya que ambos tests son equivalentes.

Ejercicio 173 *Se desea comparar el funcionamiento de cuatro dispositivos eléctricos A, B, C, D bajo distintos niveles de tensión* T_1, T_2, T_3. *Los niveles de eficiencia son:*

	T_1	T_2	T_3
A	*4*	*3*	*9*
B	*7*	*9*	*10*
C	*2*	*7*	*8*
D	*8*	*8*	*4*

Considerando la tensión como bloque (no hay interacción entre dispositivo y tensión), se pide:

1. *Construir la tabla de análisis de la varianza*
2. *Decir si hay diferencia significativa entre las eficiencias de estos dispositivos*
3. *¿Esta justificado considerar el factor tensión como bloque?*

	T_1	T_2	T_3	$\overline{y_{i\bullet}}$
A	4	3	9	5.333
B	7	9	10	8.667
C	2	7	8	5.667
D	8	8	4	6.667
$\overline{y_{\bullet j}}$	5.25	6.75	7.75	6.58

La tabla de análisis de la varianza es

FUENTE	Sum. de Cuad.	g.l..	Medias Cuad..	F$_{EXP}$	F$_{TEOR}$
Dispositivos	20.25	3	6.75	0.92	$F_{3,6} = 4.76$
Tensiones	12.6667	2	6.3334	0.86	$F_{2,6} = 5.14$
Residual	44	6	7.3333		
Total	76.9167	11			

Se deduce que no hay diferencia entre la eficiencia de los dispositivos ya que $0.92 \leq F^{-1}_{F^{-1}_{2,6}}(0.95) = 4.76$

No es razonable considerar como bloque la tensión, ya que no tiene influencia en la eficiencia de los dispositivos puesto que
$0.86 \leq F^{-1}_{F^{-1}_{3,6}}(0.95) = 5.14.$

Ejercicio 174 *Tres agentes inmobiliarios fueron interrogados acerca del precio de cinco viviendas de un barrio. Las valoraciones de estas viviendas segun los agentes se dan a continuación:*

	Agente A	*Agente B*	*Agente C*
vivienda 1	*210*	*218*	*226*
vivienda 2	*192*	*190*	*198*
vivienda 3	*183*	*187*	*185*
vivienda 4	*227*	*223*	*237*
vivienda 5	*242*	*240*	*237*

1. *Construir la tabla de análisis de la varianza*
2. *Contrastar la igualdad entre las valoraciones medias de las viviendas.*
3. *Intervalo de confianza para el valor medio de la vivienda 5.*

1)
Los valores medios se dan en la siguiente tabla

	Agente A	Agente B	Agente C	
vivienda 1	210	218	226	218
vivienda 2	192	190	198	193,333
vivienda 3	183	187	185	185,0
vivienda 4	227	223	237	229,0
vivienda 5	242	240	237	239,667
	210,8	211,6	216,6	213

1. Consideramos el modelo de dos factores sin interacción:

$$STC = SCF_A + SCF_B + SCR$$

$STC = (210 - 213)^2 + (218 - 213)^2 + (226 - 213)^2 + (192 - 213)^2 + (190 - 213)^2 + (198 - 213)^2$

$+(183 - 213)^2 + (187 - 213)^2 + (185 - 213)^2 + (227 - 213)^2 + (223 - 213)^2 + (237 - 213)^2$

$+(242 - 213)^2 + (240 - 213)^2 + (237 - 213)^2 = 6776.0$

$SCFviv = 3\,(218 - 213)^2 + 3\,(193,333 - 213)^2 + 3\,(185,0 - 213)^2 + 3\,(229,0 - 213)^2 +$

$3\,(239,667-213)^2 = 75.0+1160.\,4+2352.0+768.0+2133.\,4 = 6488.\,8$

$SCF_{AGE} = 5(210.8-213)^2+5(211.6-213)^2+5(216.6-213)^2 = 98.\,8$

$SCR = 6776.0 - 6488.\,8 - 98.\,8 = 188.\,4$

FUENTE	Sumas de Cuad	g.l.	Medias de Cuad	F_{exp}	F_{teor}
Viviendas	6488.8	4	1622. 2	$\frac{1622.\,2}{23.\,55} = 68.88$	$F_{4,8} = 3.84$
Agentes	98. 8	2	49. 4	$\frac{49.\,4}{23.\,55} = 2.097$	$F_{2,8} = 4.46$
Residual	188.4	8	23. 55		
Total	6776.0	14			

2)

Contraste para las viviendas:

La F teórica para $\alpha = 0.05$, $F_{4,8}(0.05) = 3.84 < 68.883$. Se rechaza la hipótesis de igualdad de las medias en el precio de las viviendas

Contraste para los agentes:

La F teórica para $\alpha = 0.05$, $F_{2,8}(0.05) = 4.46 > 2.\,097\,7$. No hay suficiente evidencia para rechazar la igualdad en la calificación media de los agentes. Se acepta por lo tanto la hipótesis de igualdad entre los agentes. .

3)

Un intervalo de confianza al 95% para la media de la vivienda 5 es

$\left(239.667 - t_8(0.025)\frac{\sqrt{23.55}}{\sqrt{3}}, 239.667 + t_8(0.025)\frac{\sqrt{23.55}}{\sqrt{3}}\right) =$

$\left(239.667 - 2.306\frac{\sqrt{23.55}}{\sqrt{3}}, 239.667 + 2.306\frac{\sqrt{23.55}}{\sqrt{3}}\right) = (233.\,21, 246.\,13)$

Ejercicio 175 *Un ingeniero que ha de diseñar una batería ha probado su duración para distintos materiales y soportando distintas temperaturas. La duración en horas viene dada en la tabla siguiente:*

Tipo de Material	*Temperatura en grados Farenheit*		
	15º	*70º*	*125º*
M1	150, 188 159, 126	136, 122 106, 115	25, 70 58, 45
M2	130, 74 155, 80	34, 80 40, 75	20, 82 70, 58
M3	138, 110 168, 160	174, 120 150, 139	96, 104 82, 60

1. *Hallar la suma de total de cuadrados, la suma de cuadrados correspondientes al factor temperatura, al factor material y a la interacción de ambos factores.*

2. *Indica si son significativos los efectos de la temperatura, del material y de la interacción entre ambos factores sobre la duración de las baterías.*

3. *¿Hay alguna combinación de temperatura y material que sea significativamente mayor o significativamente menor?*

1. Tabla de medias

T. M	Temperatura en grados Farenheit			
	15^o	70^o	125^o	
M1	155.75	119.75	49.5	108.333
M2	109.75	57.25	57.5	74.8333
M3	144.0	145.75	85.5	125.083
	136.5	107.583	64.1667	102.75

$STC = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} (y_{ijk} - \overline{y})^2 = 72474.8$

$SCF_{mat.} = \sum_{i=1}^{a} n_{i\cdot} (\overline{y}_{i\cdot} - \overline{y})^2 =$

$12{\times}(108.33 - 102.75)^2{+}12{\times}(74.833 - 102.75)^2{+}12{\times}(125.083 - 102.75)^2 = 15711.3$

$SCF_{temp.} = \sum_{j=1}^{b} n_{\cdot j} \left(\overline{y}_{\cdot j} - \overline{y}\right)^2 =$

$12{\times}(136.5 - 102.75)^2{+}12{\times}(107.583 - 102.75)^2{+}12{\times}(64.1667 - 102.75)^2$
$= 31813.1$

$SCF_{inter.} = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \left(\overline{y}_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y}\right)^2$

$= 4 \times (155.75 - 108.333 - 136.5 + 102.75)^2 +$

$+4 \times (119.75 - 108.333 - 107.583 + 102.75)^2$

$+4 \times (49.5 - 108.333 - 64.1667 + 102.75)^2 +$

$= 4 \times (109.75 - 74.8333 - 136.5 + 102.75)^2 +$

$+4 \times (57.25 - 74.8333 - 107.583 + 102.75)^2 +$

$+4 \times (57.5 - 74.8333 - 64.1667 + 102.75)^2 +$

$= 4 \times (144 - 125.083 - 136.5 + 102.75)^2 +$

$+4 \times (145.75 - 125.083 - 107.583 + 102.75)^2 +$

$+4 \times (85.5 - 125.083 - 64.1667 + 102.75)^2$

$= 2560.75 + 3821.66 + 1886.93 = 8269.34$

Por lo tanto la suma de los cuadrados de los residuales es

$SCR = 72474.8 - 15711.3 - 31813.1 - 8269.34 = 16681.1$

2. Las medias cuadráticas valen

$MCF_{mat} = \frac{15711.3}{3-1} = 7855.65$

$MCF_{temp.} = \frac{31813.1}{3-1} = 15906.6$

$MCF_{inter.} = \frac{8269.34}{9-3-3+1} = 2067.34$

$MCR = \frac{16681.1}{36-3\times 3} = 617.819$

F.de Var.	*S.dec.*	*G.L.*	*Medias cuad.*	*EstadísticoF*
temperatura	31813.167	2	15906.6	$\frac{15906.6}{617.819} = 25.7464$
material	15711.500	2	7855.65	$\frac{7855.65}{617.819} = 12.7151$
Interacción	8269.3333	4	2067.34	$\frac{2067.34}{617.819} = 3.34619$
RESIDUAL	16680.75027	27	617.819	
TOTAL	72474.8	35		

El valor de F correspondiente a la interacción 3.34619, $F_{4,27}(3.34619) =$ FDist$(3.34619; 4, 27) = 0.97613 > 0.95$ está en la zona de rechazo del test, por lo que no procede aceptar la hipótesis de ausencia de interacción. Se concluye que hay interacción entre los factores.

Otra forma de verlo es comparar los valores de F:

FInv$(0.95; 4, 27) = 2.72777 < 3.34619$. En este caso el estudio sobre la igualdad entre los niveles de cada factor carece de sentido, por lo que sería adecuado analizar la igualdad de las 9 combinaciones de los niveles de los factores.

3. Realizando el análisis de varianza con un solo factor con 9 niveles,cada uno de ellos con 6 replicas se concluye, como era de esperar, que las medias de las casillas no coinciden. Realizando contrastes multiples (método de Bomferroni) para las diferencias de las medias de la casilla más altas, las combinaciones de 15 grados-material 1 y 70 grados-material 3, dan el siguiente intervalo para la diferencias de las medias $(10 - 62.62, 10 + 62.2) = (-52.62,\ 72.62)$. Como este intervalo contiene el valor 0, se deduce que no hay diferencia significativa entre los valores más altos. Igualmente considerando las medias más bajas: 49.5 y 57.25 se obtiene como intervalo de confianza $(-7.75 - 62.62, -7.75 + 62.2) = (-70.37, 54.45)$, que tambien contiene

el valor 0. Por tanto no puede decirse que haya una combinación de temperatura y material que sea significativamente mejor que las restante ni tampoco ninguna combinación que sea peor que el resto.

Unidad Temática V

PROBLEMAS DE ANÁLISIS MULTIVARIANTE

T. 15

Análisis multivariante. Regresión

Ejercicio 176 *Los tabla siguiente indica la edad, los años de experiencia y los ingresos mensuales (en miles de pesetas) de 5 ingenieros.*

Edad	*37*	*45*	*38*	*42*	*31*
Experiencia	*4*	*0*	*5*	*2*	*4*
Ingresos	*512*	*468*	*550*	*503*	*454*

El modelo de regresión lineal que relacione los ingresos con las otras dos variables es:

ingresos = 37.21 + 9.61 edad + 29.76 experiencia

1. *Calcular el coeficiente de determinación*
2. *¿Es la regresión significativa?*
3. *Emplea este ajuste para predecir cuanto ganan por promedio los ingenieros de 40 años de edad y 4 de experiencia*

1. Valores estimados

$$\left(\widehat{yj}\right) = 37.21 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + 9.61 \begin{pmatrix} 37 \\ 45 \\ 38 \\ 42 \\ 31 \end{pmatrix} + 29.76 \begin{pmatrix} 4 \\ 0 \\ 5 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 511.82 \\ 469.66 \\ 551.19 \\ 500.35 \\ 454.16 \end{pmatrix}$$

$$\left(\widehat{yj} - \bar{y}\right) = \begin{pmatrix} 511.82 \\ 469.66 \\ 551.19 \\ 500.35 \\ 454.16 \end{pmatrix} - \begin{pmatrix} 497.4 \\ 497.4 \\ 497.4 \\ 497.4 \\ 497.4 \end{pmatrix} = \begin{pmatrix} 14.42 \\ -27.74 \\ 53.79 \\ 2.95 \\ -43.24 \end{pmatrix}$$

$$(yj-\bar{y}) = \begin{pmatrix} 512 \\ 468 \\ 550 \\ 503 \\ 454 \end{pmatrix} - \begin{pmatrix} 497.4 \\ 497.4 \\ 497.4 \\ 497.4 \\ 497.4 \end{pmatrix} = \begin{pmatrix} 14.6 \\ -29.4 \\ 52.6 \\ 5.6 \\ -43.4 \end{pmatrix}$$

$$(yj-\widehat{y_j}) = \begin{pmatrix} 512 \\ 468 \\ 550 \\ 503 \\ 454 \end{pmatrix} - \begin{pmatrix} 511.82 \\ 469.66 \\ 551.19 \\ 500.35 \\ 454.16 \end{pmatrix} = \begin{pmatrix} .18 \\ -1.66 \\ -1.19 \\ 2.65 \\ -.16 \end{pmatrix}$$

$$R^2 = \frac{\sum\limits_{j=1}^{n}\left(\widehat{yj}-\bar{y}\right)^2}{\sum\limits_{j=1}^{n}(yj-\bar{y})^2} = \frac{14.42^2+27.74^2+53.79^2+2.95^2+43.24^2}{14.6^2+29.4^2+52.6^2+5.6^2+43.4^2} = \frac{5749.2}{5759.2} = .99827$$

2)

$$F = \frac{\text{media de cuadrados explicados por la regresión}}{\text{media de cuadrados no explicados por la regresión}} = \frac{\frac{\sum_{j=1}^{n}\left(\widehat{yj}-\bar{y}\right)^2}{r}}{\frac{\sum_{j=1}^{n}(yj-\widehat{y_j})^2}{n-r-1}} =$$

$$= \frac{\frac{14.42^2+27.74^2+53.79^2+2.95^2+43.24^2}{2}}{\frac{.18^2+1.66^2+1.19^2+2.65^2+.16^2}{2}} = 510.94$$

$\mathrm{FInv}(0.95;2,2) = 19.0 < 510.94$.

Por lo tanto se rechaza la hipótesis nula, y se concluye que la regresión es significativa

3)

$37.21 + 9.61 \times 40 + 29.76 \times 4 = 540.65$ *Miles de Pesetas.*

Ejercicio 177 *En un estudio de consumo se estimo la siguente ecuación, obtenida con 200 datos:*

$\log(y) = -0.243 - 0.562\log x_1 + 0.327\log x_2 + 0.219\log x_3 - 0.127\log x_4$, *siendo el coeficiente de determinación del ajuste* $R^2 = 0.853$, *y 0.219, 0.161, 0.157, 0.082 los errores estandar correspondientes a los coeficientes del ajuste.*

$y =$ *Cantidad de carne de cerdo comprada*

$x_1 =$ *Precio de la carne de cerdo,* $x_2 =$ *Precio de la carne de ternera,* $x_3 =$ *Precio de la carne de pollo,* $x_4 =$ *ingreso medio por familia.*

1. *Interpretar los coeficientes del modelo de regresión*

2. *Indicar que variables son significativas.*

1. El modelo ajustado

   $\log(y) = -0.243 - 0.562 \log x_1 + 0.327 \log x_2 + 0.219 \log x_3 - 0.127 \log x_4,$

   es equivalente a:

   $y = 0.784\,27 \dfrac{x_2^{0.327} x_3^{0.219}}{x_1^{0.562} x_4^{0.127}}.$

   La cantidad de carne de cerdo comprada, y, aumenta , si aumenta el precio de la carne de ternera y de pollo (variables de coeficiente positivo, x_2 y x_3), y disminuye si aumenta el precio de la carne de cerdo y los ingresos medios familiares (x_1 y x_4 variables con coeficiente negativo en la regresión).

   Así que si, por ejemplo, la carne de ternera, x_2, aumenta al doble, la cantidad de carne de cerdo comprada, y, se espera que aumente en la proporción: $2^{0.327} = 1.\,254\,4$.

   Hallando los intervalos de confianza para los parámetros de la recta regresión se aceptan como parámetros significativos aquellos cuyos intervalos de confianza no contengan el valor 0.

   $(0.562 - t_{200-3}^{-1}(0.975)0.219, 0.562 + t_{200-3}^{-1}(0.975)0.219)$

   sustituyendo $t_{197}^{-1}(0.975) = 1.9271$ resulta el intervalo:

   $(0.562 - 1.9721 \times 0.219, 0.562 + 1.9721 \times 0.219) = (0.130\,11, 0.993\,89).$
   Por tanto este coeficiente es significativo.

   Este cálculo es equivalente a realizar estos test de hipótesis:

$\frac{0.562}{0.219} = 2.\,566\,2 > 1.9271$ El primer coeficiente es significativo.
$\frac{0.327}{0.161} = 2.\,031\,1 > 1.9721$, El segundo coeficiente es significativo.
$\frac{0.219}{0.157} = 1.\,394\,9 < 1.\,972\,1$ El tercer coeficiente no es significativo
$\frac{0.127}{0.082} = 1.\,548\,8 < 1.9721$. El cuarto coeficiente no es significativo.
No se acepta que el log de la cantidad de carne comprada dependa linealmente de log de la carne de pollo ni del log del ingreso medio por familia. En este caso sería conveniente eliminar estas variables del modelo de regresión .

Ejercicio 178 *Para 11 provincias españolas se conocen los siguientes datos:*
Y =número de mujeres conductoras dividido por el número de hombres conductores
X_1 = Porcentaje de mujeres trabajadoras sobre el total de trabajadores de la provincia
X_2 = Porcentaje de población que trabaja en el sector agrícola.
para obtener el modelo de regresión lineal donde la primera variable es dependiente y las dos restantes independientes, se ha obtenido:

$$(X'X)^{-1} = \begin{pmatrix} 5.1 & -0.12 & -0.05 \\ -0.12 & 30.8 & 0.08 \\ -0.05 & 0.08 & 0.001 \end{pmatrix}, \quad (X'Y) = \begin{pmatrix} -0.06 \\ 0.05 \\ -9.45 \end{pmatrix}$$

$\widehat{\sigma^2} = 0.003; \quad \sum (y - \overline{y})^2 = 0.0645$

1. Estimar el modelo de regresión y los contrastes individuales para los coeficientes.

2. Calcular el coeficiente de determinación.

1. $\hat{\mathbf{b}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} =$

$$= \begin{pmatrix} 5.1 & -0.12 & -0.05 \\ -0.12 & 30.8 & 0.08 \\ -0.05 & 0.08 & 0.001 \end{pmatrix} \begin{pmatrix} -0.06 \\ 0.05 \\ -9.45 \end{pmatrix} = \begin{pmatrix} 0.16 \\ 0.791 \\ -0.0025 \end{pmatrix}$$

$Y = 0.1605 + 0.7912X_1 - 0.00245X_2$

Los errores estandar de los coeficientes son:

$\sqrt{\widehat{\sigma}^2 C_{00}} = \sqrt{0.003 \times 5.1} = 0.123\,69$

$\sqrt{\widehat{\sigma}^2 C_{11}} = \sqrt{0.003 \times 30.8} = 0.303\,97$

$\sqrt{\widehat{\sigma}^2 C_{22}} = \sqrt{0.003 \times 0.001} = 1.\,732\,1 \times 10^{-3}$

Realizamos los test de hipótesis de la t de Student correspondiente a la significación de los coeficientes:

$t^{-1}_{11-2-1}(0.975) = t^{-1}_{8}(0.975) = 2.\,306$

$\frac{0.160\,5}{0.123\,69} = 1.\,297\,6 < 2.306$. No es significativo

$\frac{0.791\,2}{0.303\,97} = 2.\,602\,9 > 2.306$. Es significativo

$\frac{0.002\,45}{1.\,732\,1 \times 10^{-3}} = 1.\,414\,5 < 0.306$. No es significativo

2. $R^2 = \frac{\sum_{j=1}^{n} (\widehat{y}j - \overline{y})^2}{\sum_{j=1}^{n} (yj - \overline{y})^2}$

Como

$$\sum_{j=1}^{n} (y_j - \overline{y})^2 = \sum_{j=1}^{n} (\widehat{y_j} - \overline{y})^2 + \sum_{j=1}^{n} (y_j - \widehat{y_j})^2;$$

$$0.00645 = \sum_{j=1}^{n} (\widehat{y_j} - \overline{y})^2 + (n - r - 1)S_R^2 = \sum_{j=1}^{n} (\widehat{y_j} - \overline{y})^2 + 8 \times 0.03$$

$$\sum_{j=1}^{n} (\widehat{y_j} - \overline{y})^2 = 0.0645 - 8 \times 0.003 = 0.040\,5$$

$$R^2 = \frac{\sum_{j=1}^{n} (\widehat{y}j - \overline{y})^2}{\sum_{j=1}^{n} (yj - \overline{y})^2} = \frac{0.040\,5}{0.0645} = 0.627\,91$$

Ejercicio 179 *Con los datos de los 12 meses del año 1973 de la encuesta de presupuestos familiares se han probado seis distintos modelos de regresión lineal (sin constante) en los que la variable dependiente es $GTINE$ (Gasto Total según el INE) y las variables explicativas son las siguientes:*

$IT =$ *Ingreso Total*

$G6 =$ *Gasto en transporte y comunicaciones*

$G7 =$ *Gasto en esparcimiento y enseñanza.*

Los coeficientes de los modelos estudiados y sus errores estándar(entre paréntisis se indican en la tabla siguiente:

	M1	*M2*	*M3*	*M4*	*M5*	*M6*
IT	*0.79* *(0.09)*			*0.69* *(0.11)*	*0.59* *(0.08)*	*0.57* *(0.09)*
G6		*3.04* *(0.59)*		*0.79* *(0.62)*		*0.18* *(0.50)*
G7			*3.33* *(0.43)*		*2.35* *(0.35)*	*2.33* *(0.36)*
R^2	*49.40*	*26.16*	*44.49*	*50.53*	*68.61*	*68.67*

Comentar los resultados y elegir el mejor modelo.

Se han probado modelos de una, dos y tres variables explicativas. Conforme se introduce una nueva variable explicativa aumenta el coeficiente de determinación R^2. Aunque el ajuste el error cuadrático medio sea relativamente más pequeño no es conveniente que en el modelo haya variables no significativas:

M1) $\frac{0.79}{0.09} = 8.7778 > t_{10}^{-1}(0.975) = 2.2281$ Significativa.
M2) $\frac{3.04}{0.59} = 5.1525 > t_{10}^{-1}(0.975; 10) = 2.2281$ Significativa.
M3) $\frac{3.33}{0.43} = 7.7442 > t_{10}^{-1}(0.975; 10) = 2.2281$ Significativa.
M4)

$\frac{0.69}{0.11} = 6.2727 > t_9^{-1}(0.975; 9) = 2.2622$ Significativa
$\frac{0.79}{0.62} = 1.2742 < t_9^{-1}(0.975; 9) = 2.2622$ No Significativa

M5)

$\frac{0.59}{0.08} = 7.375 > t_9^{-1}(0.975; 9) = 2.2622$ Significativa
$\frac{2.35}{0.35} = 6.7143 > t_9^{-1}(0.975; 9) = 2.2622$ Significativa

M6)

$\frac{0.57}{0.09} = 6.3333 > t_{10}^{-1}(0.975; 8) = 2.306$ Significativa
$\frac{0.18}{0.50} = 0.36 < t_{8}^{-1}(0.975; 8) = 2.306$ No significativa
$\frac{2.33}{0.36} = 6.4722 > t_{8}^{-1}(0.975; 8) = 2.306$ Significativa

Entre los modelos con todas las variables explicativas significativas (1, 2, 3, 5) elijo el de mayor coeficiente de determinación, que es el modelo 5.

$$GTINE = 0.59 \times IT + 2.35 \times G7$$

Ejercicio 180 *Los siguientes datos se refieren a seis pisos que pone a la venta una agencia inmobiliaria. Los datos son un índice de valoración del barrio en el que se ubica, la distancia desde cada piso al centro escolar más próximo en km y el precio por metro cuadrado de la vivienda. Estudia el modelo de regresión lineal multiple cuya variable dependiente es el precio y las otras dos las variables independientes o regresoras.*

barrio	*Distancia*	*precio*
4	*1.5*	*1600*
3	*2.2*	*1120*
1.6	*1.0*	*690*
1.2	*2.0*	*900*
3.4	*0.8*	*1230*
4.8	*1.6*	*1860*

1) Ajuste del modelo:
Calculamos en primer lugar el ajuste de regresión:
$\widehat{\mathbf{b}} = (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}\prime\mathbf{y}$

$$(\mathbf{x}'\mathbf{x}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 3 & 1.6 & 1.2 & 3.4 & 4.8 \\ 1.5 & 2.2 & 1 & 2 & 0.8 & 1.6 \end{pmatrix} \begin{pmatrix} 1 & 4 & 1.5 \\ 1 & 3 & 2.2 \\ 1 & 1.6 & 1 \\ 1 & 1.2 & 2 \\ 1 & 3.4 & 0.8 \\ 1 & 4.8 & 1.6 \end{pmatrix} =$$

$$= \begin{pmatrix} 6.0 & 18.0 & 9.1 \\ 18.0 & 63.6 & 27.0 \\ 9.1 & 27.0 & 15.29 \end{pmatrix}$$
,

$$(\mathbf{x}\prime\mathbf{x})^{-\mathbf{1}} = \begin{pmatrix} 2.8577 & -0.34653 & -1.0889 \\ -0.34653 & 0.10483 & 2.1130 \times 10^{-2} \\ -1.0889 & 2.1130 \times 10^{-2} & 0.67615 \end{pmatrix}$$

$\hat{\mathbf{b}} = (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}\prime\mathbf{y} =$

$$= \begin{pmatrix} 2.8577 & -0.34653 & -1.0889 \\ -0.34653 & 0.10483 & 2.1130\times10^{-2} \\ -1.0889 & 2.1130\times10^{-2} & 0.67615 \end{pmatrix} \times$$

$$\times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 3 & 1.6 & 1.2 & 3.4 & 4.8 \\ 1.5 & 2.2 & 1 & 2 & 0.8 & 1.6 \end{pmatrix} \times \begin{pmatrix} 1600 \\ 1120 \\ 690 \\ 900 \\ 1230 \\ 1860 \end{pmatrix} = \begin{pmatrix} 145.2 \\ 301.15 \\ 121.49 \end{pmatrix}$$

$=$

Así que el plano de regresión es en este caso

$$y = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} 145.2 \\ 301.15 \\ 121.49 \end{pmatrix} = 145.2 + 301.15x_1 + 121.49x_2$$

La relación obtenida es Precio = 145.2 + 301.15 $\times$ barrio + 121.49$\times$ Distancia

2) prueba de la significación de la regresión:

Realizamos ahora el test de hipótesis de significación de la regresión. Para ello evaluamos para cada punto los valores predichos por la regresión:

x_1	4	3	1.6	1.2	3.4	4.8
x_2	1.5	2.2	1	2	0.8	1.6
y	1600	1120	690	900	1230	1860
$\widehat{y}$	1532	1315.9	748.53	749.56	1266.3	1785.1

Media de y = 1233.3

$$F = \frac{\dfrac{\sum_{j=1}^{n}\left(\widehat{yj} - \bar{y}\right)^2}{r}}{\dfrac{\sum_{j=1}^{n}\left(yj - \widehat{y_j}\right)^2}{n-r-1}};$$

$\sum_{j=1}^{6}\left(\widehat{yj} - \bar{y}\right)^2 = (1532 - 1233.3)^2 + (1315.9 - 1233.3)^2 + (748.53 - 1233.3)^2 + (749.56 - 1233.3)^2 + (1266.3 - 1233.3)^2 + (1785.1 - 1233.3)^2 = 8.7062 \times 10^5$

$\sum_{j=1}^{6}\left(yj - \widehat{y_j}\right)^2 = (1600 - 1532)^2 + (1120 - 1315.9)^2 + (690 - 748.53)^2 + (900 - 749.56)^2 + (1230 - 1266.3)^2 + (1860 - 1785.1)^2 = 75986.$

Por tanto

$$F = \frac{\dfrac{\sum_{j=1}^{n}\left(\widehat{yj} - \bar{y}\right)^2}{r}}{\dfrac{\sum_{j=1}^{n}\left(yj - \widehat{y}\right)^2}{n-r-1}} = \frac{\frac{8.7062\times10^5}{2}}{\frac{75986}{3}} = 17.186$$

Como la $F_{2,3}^{-1}(0.95) = 9.55209$ que es mayor que la experimental se rechaza la hipótesis nula al nivel 0.05. Como la hipótesis nula de este test es quen los coeficientes son nulos, se concluye que al menos alguno de los coeficientes del ajuste no es nulo, y por tanto existe relación lineal entre la variable PRECIO y al menos una de las variables BARRIO y DISTANCIA. Si hallamos el P_VALUE del test: $p-value = 1\text{-}F_{2,3}(17.186; 2, 3) = 1-0.977\,26 = 0.022\,74$. Los valores menores que el nivel de significación del test $(0.0274 < 0.05)$ nos indican que se debe rechazar la hipótesis nula, lo que por supuesto nos lleva a la misma conclusión

3) Coeficiente de determinación

$R^2 = \frac{8.\,706\,2\times10^5}{75986+8.\,706\,2\times10^5} = 0.919\,73$

El estadístico R-cuadrado ajustado indica que el 91.7% de la variabilidad del precio está explicado por el modelo dado de regresión lineal multiple.

4) Intervalos de confianza para los parámetros del ajuste.

En primer lugar calculamos la estimación de la varianza:

$$\widehat{\sigma}^2 = \frac{\sum_{j=1}^{n} (yj - \widehat{y})^2}{n-r-1} = \frac{75986}{3} = 25329.$$

Los intervalos de confianza para los coeficientes son

$+145.\,2 + 301.\,15x_1 + 121.\,49x_2$

$$\begin{pmatrix} 2.\,857\,7 & -0.346\,53 & -1.\,088\,9 \\ -0.346\,53 & 0.104\,83 & 2.\,113\,0\times10^{-2} \\ -1.\,088\,9 & 2.\,113\,0\times10^{-2} & 0.676\,15 \end{pmatrix}$$

$\left(\widehat{b}_0 - t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{00}} < b_0 < \widehat{b}_0 + t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{00}}\right) =$

$= \left(145.2 - 4.3027\sqrt{25329\times2.8577}, 145.2 + 4.3027\sqrt{25329\times2.8577}\right) =$

$= (-1012.\,4, 1302.\,8)$

$\left(\widehat{b}_1 - t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{11}} < b_1 < \widehat{b}_1 + t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{11}}\right) =$

$=\left(301.15 - 4.3027\sqrt{25329\times0.10483}, 301.15 + 4.3027\sqrt{25329\times0.10483}\right) =$ $(79.\,436, 522.\,86)$

$\left(\widehat{b}_2 - t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{22j}} < b_2 < \widehat{b}_2 + t_{n-r-1}(\frac{\alpha}{2})\sqrt{\widehat{\sigma}^2 C_{22}}\right) =$

$= \left(121.\,49 - 4.3027\sqrt{25329\times0.676\,15}, 121.\,49 + 4.3027\sqrt{25329\times0.676\,15}\right) =$ $(-441.\,59, 684.\,57)$

Solo es significativo el coeficiente de BARRIO, ya que los otros intervalos de confianza contienen el valor 0. Esto nos sugiere que es más apropiado usar el modelo lineal $PRECIO = b \times BARRIO$.

5) Cálculo del modelo $PRECIO = b_1 \times BARRIO$ **:**

Empleamos en esta ocasión el método de los mínimos cuadrados a los datos:

x_1 = barrio	y = precio
4	1600
3	1120
1.6	690
1.2	900
3.4	1230
4.8	1860

La suma de los cuadrados de los residuales es $\sum\limits_{i=1,6} (y_1 - b_1 x_{i1})^2 =$
$(1600 - 4b_1)^2 + (1120 - 3b_1)^2 + (690 - 1.6b_1)^2 +$
$+ (900 - 1.2b_1)^2 + (1230 - 3.4b_1)^2 + (1860 - 4.8b_1)^2 =$
$= 63.\,6b_1^2 - 50108.b_1 + 10\,073\,000$
Derivando, igualando a 0 y resolviendo la ecuación calculo el valor del nuevo coeficiente b_1 :
$127.2b_1 = 50108$
La solución es $b_1 = 393.\,93$
por lo tanto el ajuste es: $PRECIO = 393.93 \times BARRIO$
El coeficiente de determinación se define como:
$R^2 = 1 - \frac{\sum_{j=1}^{n}(yj - \widehat{y_j})^2}{\sum_{j=1}^{n}(yj - \overline{y_j})^2} = 1 - \frac{2.\,03457}{9\,463\,33} = 0.785\,.$
Este ajuste tiene más precisión en los coeficientes, aunque la bondad de ajuste a los puntos es menor.

Ejercicio 181 *Dados los puntos de coordenadas (20,22) (16,41) (10,120) (11,89) (14,56) ajusta una parabola a estos puntos considerando la segunda variable como variable dependiente usando los procedimientos siguientes:*

1. *Utilizando las ecuaciónes normales de ajuste polinomial.*
2. *Usando un modelo de regresión multiple y el enfoque matricial.*
3. *Usando algún paquete estadístico.*

1. El procedimiento consiste en hallar los parámetros del ajuste $y = ax^2 + bx + c$ que minimice la varaianza residual, o equivalentemente que minimice la suma de los cuadrados de los residuales:

   $\sum\limits_{i=1}^{n} (y_i - \left(ax^2 + bx + c\right))^2 =$

   $\sum\limits_{i=1}^{n} y_i^2 + \sum\limits_{i=1}^{n} \left(ax^2 + bx + c\right)^2 - 2\sum\limits_{i=1}^{n} y_i \left(ax^2 + bx + c\right)$

Derivando con respecto a los parámetros $(a,\ b,\ c)$, e igualando a 0 se obtiene el siguiente sistema:

$a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x^2$

$a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_1 = \sum_{i=1}^{n} y_i x_i$

$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x + cn = \sum_{i=1}^{n} y_i$

La siguiente tabla recoge los cálculos intermedios necesario para hallar los coeficientes del sistema:

x	x^2	x^3	x^4	y	yx	yx^2
20	400	8000	160000	22	440	8800
16	256	4096	65536	41	656	10496
10	100	1000	10000	120	1200	12000
11	121	1331	14641	89	979	10769
14	196	2744	38416	56	784	10976
71	**1073**	**17171**	**288592**	**328**	**4059**	**53041**

El sistema de las ecuaciones normales resulta:

$$\begin{gathered} 288593a + 17171b + 1073c = 53041 \\ 17171a + 1073b + 71c = 4059 \\ 1073a + 71b + 5c = 328 \end{gathered}$$

Resolviendo el sistema obtenemos los valores de los coeficientes $a,\ b,\ c:$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 288598 & 17171 & 1073 \\ 17171 & 1073 & 71 \\ 1073 & 71 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 53041 \\ 4059 \\ 328 \end{pmatrix} = \begin{pmatrix} 0.888\,78 \\ -35.\,769 \\ 382.\,79 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{9}{5281} & -\frac{806}{15\,843} & \frac{5651}{15\,843} \\ -\frac{806}{15\,843} & \frac{291\,661}{190\,116} & -\frac{2065\,975}{190\,116} \\ \frac{5651}{15\,843} & -\frac{2065\,975}{190\,116} & \frac{14\,822\,413}{190\,116} \end{pmatrix} \begin{pmatrix} 53041 \\ 4059 \\ 328 \end{pmatrix} = \begin{pmatrix} 0.888\,78 \\ -35.769 \\ 382.\,79 \end{pmatrix}$$

$a = 0.888\,78,\ b = -35.\,769, .c = 382.739$

De modo que el ajuste resulta:

$y = 0.888\,78\ x^2 - 35.\,769\ x + 382.\,79$

2. También se puede realizar un ajuste parabólico usando el procedimiento de regresión multiple considerando como variables independientes $x_1 =$

x, $x_2 = x^2$

$x_1 = x$	$x_2 = x^2$	y
20	400	22
16	256	41
10	100	120
11	121	89
14	196	56

Empleando el método matricial: $\widehat{\mathbf{b}}= (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}\prime\mathbf{y}$

$$(\mathbf{x}'\mathbf{x}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 16 & 10 & 11 & 14 \\ 400 & 256 & 100 & 121 & 196 \end{pmatrix} \begin{pmatrix} 1 & 20 & 400 \\ 1 & 16 & 256 \\ 1 & 10 & 100 \\ 1 & 11 & 121 \\ 1 & 14 & 196 \end{pmatrix} =$$

$$= \begin{pmatrix} 5 & 71 & 1073 \\ 71 & 1073 & 17\,171 \\ 1073 & 17\,171 & 288\,593 \end{pmatrix}$$

$$\mathbf{x}\prime\mathbf{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 16 & 10 & 11 & 14 \\ 400 & 256 & 100 & 121 & 196 \end{pmatrix} \begin{pmatrix} 22 \\ 41 \\ 120 \\ 89 \\ 56 \end{pmatrix} = \begin{pmatrix} 328.0 \\ 4059.0 \\ 53041. \end{pmatrix}$$

$$\widehat{\mathbf{b}}= (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}\prime\mathbf{y} = \begin{pmatrix} 5 & 71 & 1073 \\ 71 & 1073 & 17\,171 \\ 1073 & 17\,171 & 288\,598 \end{pmatrix}^{-1} \begin{pmatrix} 328.0 \\ 4059.0 \\ 53041. \end{pmatrix} =$$

$$= \begin{pmatrix} 382.\,79 \\ -35.\,769 \\ 0.888\,78 \end{pmatrix}$$

,

Así que el plano de regresión resulta:

$y = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} = -35.\,769x_1 + 0.88878x_2 + 382.79$

y la ecuación de la parábola.

$y = 382.79 - 35.\,769x + 0.88878x^2$

1. Con el paquete Stagraphics plus 5.0 se obtiene el siguiente resultado:

Polynomial Regression - y versus x
Polynomial Regression Analysis

Dependent variable: y

Standard T

Parameter	Estimate	Error	Statistic	EP-Value
CONSTANT	384.393	65.1637	5.89888	0.0276
x	-35.9975	9.14209	-3.93756	0.0589
x^2	0.896422	0.304718	2.94181	0.0987

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	5997.16	2	2998.58	55.51	0.0177
Residual	108.039	2	54.0197		

Total (Corr.) 6105.2 4
R-squared = 98.2304 percent
R-squared (adjusted for d.f.) = 96.4607 percent
Standard Error of Est. = 7.34981
Mean absolute error = 3.60963
Durbin-Watson statistic = 2.55791 (P=0.1368)
Lag 1 residual autocorrelation = -0.283759

Comentarios

El polinomio de ajuste es:

$y = 384.393 - 35.9975\ x + 0.896422\ x^2$

El ajuste es ligeramente diferente a los anteriores posiblemente por errores numéricos.

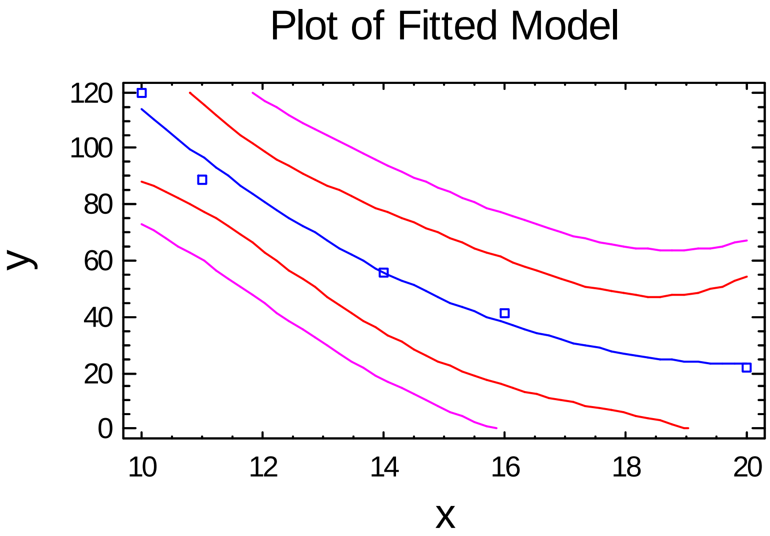
En el test de significación de la regresión resulta un $p - value = 0.0177 < 0.05$. Esto quiere decir que con una confianza del 95% se rechazrá la hipótesis nula, y por tanto hay algún tipo de dependencia, de primer o segundo grado, de y con respecto a x.

El valor del coeficiente de determinación, 0.982304 nos indica que los puntos se ajustan bastante bien a la parabola.

La estimación de la desviación típica de los residuales es 7.34981.

En cuanto a la significación de los parámetros del ajuste algunos depende del grado de confianza con el que se quiera trabajar. El menos preciso es el coeficiente de x^2, cuyo nivel de confianza sería del 90%.

La representación gráfica del ajuste, junto con los intervalos de confianza correspondiente a las predicciones para las medias de la variable dependiente para cada valor de x y tambien los intervalos de confianza correspondientes a los valores individuales de y es la de la siguiente figura:



Plot of Fitted Model
y
x
0
20
40
60
80
100
120
10
12
14
16
18
20

T. 16

Diversas técnicas de Análisis Multivariante .

16.1 Análisis de componentes principales

Ejercicio 182 *Las siguientes tablas suministran los gastos por distintos conceptos en algunas comunidades autónomas españolas (en el fichero coaut.sf3):*

AL=Alimentación, bebidas y tabaco

VES=Vestido y calzado

VIV=Vivienda, calefaccion y alumbrado

SER=Artículos de mobiliario, menaje y conservación del hogar

MED=Servicios médicos y sanitarios

TRANS=Transportes y comunicaciones

ESP=Esparcimiento, enseñanza y cultura

OBIEN=Otros bienes y servicios

OGAS=Otros gastos

COMUN	*AL*	*VES*	*VIV*	*SER*	*MED*
ANDALUCIA	*605*	*222*	*183*	*121*	*50*
ARAGON	*548*	*255*	*202*	*126*	*50*
ASTURIAS	*587*	*281*	*233*	*131*	*58*
BALEARES	*550*	*227*	*206*	*144*	*87*
CANARIAS	*572*	*186*	*180*	*134*	*74*
CANTABRIA	*588*	*289*	*261*	*118*	*64*
CASTILLA LA MANCHA	*543*	*221*	*201*	*126*	*53*
CASTILLA Y LEON	*547*	*218*	*191*	*119*	*41*
CATALUÑA	*686*	*262*	*283*	*163*	*93*
CEUTA Y MELILLA	*683*	*193*	*134*	*81*	*27*
COM. VALENCIANA	*542*	*218*	*177*	*132*	*62*
EXTREMADURA	*470*	*211*	*138*	*100*	*44*
GALICIA	*615*	*248*	*199*	*132*	*51*
LA RIOJA	*602*	*210*	*196*	*127*	*55*
MADRID	*674*	*254*	*253*	*146*	*87*
MURCIA	*604*	*210*	*189*	*128*	*47*
NAVARRA	*643*	*325*	*251*	*221*	*81*
PAIS VASCO	*636*	*267*	*232*	*158*	*65*

TRANS	*ESP*	*OBIEN*	*OGAS*	*COMUN*
255	*115*	*281*	*83*	*ANDALUCIA*
247	*111*	*263*	*82*	*ARAGON*
336	*151*	*311*	*121*	*ASTURIAS*
357	*151*	*334*	*131*	*BALEARES*
305	*159*	*280*	*98*	*CANARIAS*
302	*117*	*277*	*106*	*CANTABRIA*
244	*96*	*253*	*101*	*CASTILLA LA MANCHA*
252	*110*	*260*	*108*	*CASTILLA Y LEON*
361	*228*	*363*	*107*	*CATALUÑA*
142	*84*	*235*	*64*	*CEUTA Y MELILLA*
280	*123*	*281*	*99*	*COM. VALENCIANA*
208	*87*	*223*	*71*	*EXTREMADURA*
291	*128*	*256*	*104*	*GALICIA*
263	*126*	*313*	*122*	*LA RIOJA*
370	*216*	*433*	*130*	*MADRID*
319	*104*	*310*	*114*	*MURCIA*
407	*186*	*409*	*155*	*NAVARRA*
343	*174*	*395*	*122*	*PAIS VASCO*

Calcular con Statgraphics las tres primeras componentes principales por medio de su relación con las variables primitivas y el porcentaje de varianza *explicada por éstas componentes.*

Los coeficientes de las componentes principales respecto a las variables primitivas vienen dados en la tabla:

Componente	Autovalor	% Varianza explicada	% Varianza acumulada
1	12652.0	72.337	72.337
2	2615.62	14.955	87.291
3	994.89	5.688	92.979

La primera componente explica el 72.337 % de la variabilidad de los datos, la segunda el 14.955% y la tercera el 5.688%. Las tres componentes juntas explican el casi el 93% de la variabilidad, así que si se usan estas tres componentes en lugar de las 9 componentes primitivas sólo se pirde el 7% de la información contenida en los datos.

Estas componentes se obtiene en función de las variables primitivas por medio de una relación lineal. Los coeficientes son los siguientes:

	1^a Comp.	2^a Comp.	3^a Comp.
AL	0.304093	-0.880292	-0.115131
VES	0.226685	0.128553	-0.737529
VIV	0.307534	0.061498	-0.47601
SER	0.218531	0.135562	-0.016936
MED	0.132717	0.0829788	0.106043
$TRANS$	0.537525	0.394558	0.119099
ESP	0.342499	-0.053597	0.199809
$OBIEN$	0.512262	-0.0698458	0.384319
$OGAS$	0.165962	0.126844	0.0535598

Por ejemplo, la primera componente se expresa en función de las nueve variables primitivas de la siguiente forma:

1^a Comp.$= 0.304093 \times AL + 0.226685 \times VES + ... + 0.165963 \times OGAS$

16.2 Análisis discriminante

Ejercicio 183 *Una compañía aseguradora ha realizado una una investigación sobre la siniestralidad en los vehículos asegurados con el objeto de obtener un criterio para la admisión de nuevos clientes. Para ello ha seleccionado aleatoriamente 40 pólizas, clasificadas en dos grupos, separando a los asegurados* *que han tenido un siniestro grave de los restantes. Se desea tomar en consideración la información sobre la edad del conductor, la antigüedad del coche y su potencia. Los datos obtenidos son los del fichero siniestr.sf3. Los vehículos con siniestro grave vienen codíficados en la variable* siniestr *con un 1 y los restantes con un 2:*

1. *Aplicar la técnica de Análisis discriminante para estudiar qué información dan las tres variables sobre si un asegurado va a tener o no un siniestro grave.*
2. *Con esta información, ¿Cúal sería el pronóstico para un posible cliente de 30 años de edad, que desea asegurar un vehiculo de 5 años de antiguëdad y con valor 150 para la potencia?*

1. La salida de Statgraphics sobre el poder de previsión de las variables es:

Discriminant Function	Eigenvalue	Relative Percentage	Canonical Correlation
1	0.562838	100.00	0.60012

Functions Derived	Wilks Lambda	Chi-Square	DF	P-Value
1	0.639862	16.2974	3	0.0010

Como el p-value es menor que 0.05 las variables consideradas discriminan significativamente (con el 95% de confianza) entre los dos grupos considerados (siniestro grave o no)

El porcentaje de aciertos que nos da el programa viene resumido en la

siguiente tabla:

Classification Table

Actual SINIESTR	Group Size	Predicted SINIESTR 1	2
1	16	11 (68.75%)	5 (31.25%)
2	24	4 (16.67%)	20 (83.33%)

Percent of cases correctly classified: 77.50%

El porcentaje de datos bién clasificados es del 77.5%. De los 16 casos en que ha habido siniestro grave 11 han sido clasificado correctamente. Si hubiéramos hecho la selección al azar se esperaría acertar en sólo 8 casos. De los 24 casos en que no ha habido siniestro se ha acertado en 20. Al azar se esperaría acertar en 12 casos. Está mejoría del pronóstico del 50% al 77.50% se consigue por la información suministrada por las tres variables independientes: antiguëdad, edad y potencia.

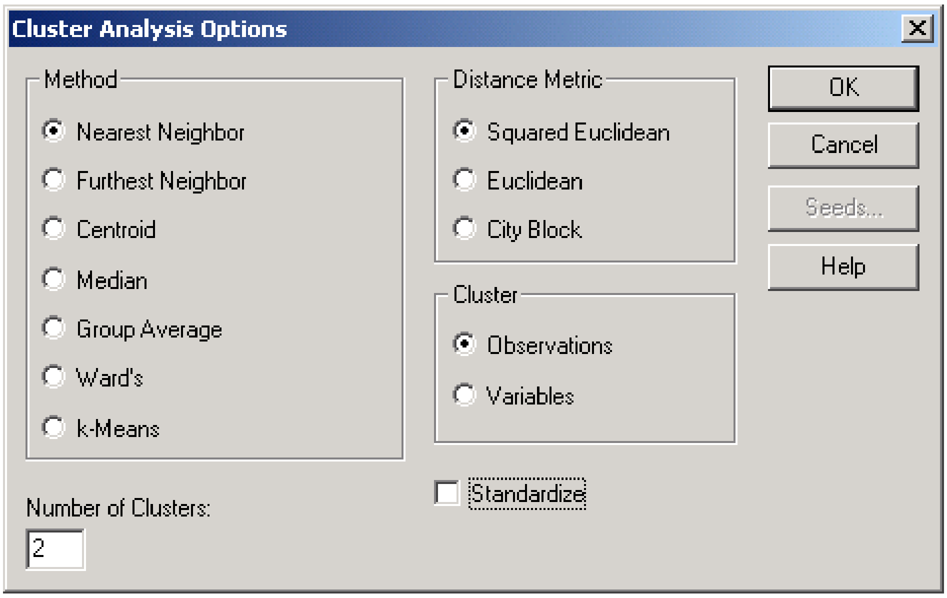
2. La previsión que hace Statgraphics para el nuevo cliente es que no va a sufrir siniestro grave.

16.3 Análisis Cluster

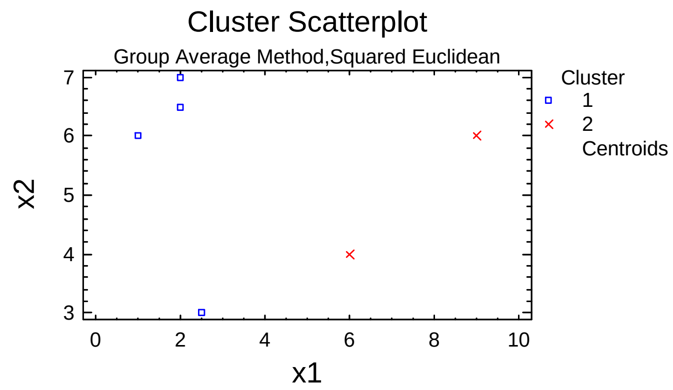
Ejercicio 184 *Se dispone de 6 observaciones y dos variables y se trata de reunir las observaciones en dos grupos, en función de su semejanza. a) Utilizar como opciones la distancia euclídea al cuadrado y para la vinculación intergrupos el método del vecino más próximo.*

Observación	*1*	*2*	*3*	*4*	*5*	*6*
x_1	*1*	*2*	*2*	*2.5*	*6*	*9*
x_2	*6*	*6.5*	*7*	*3*	*4*	*6*

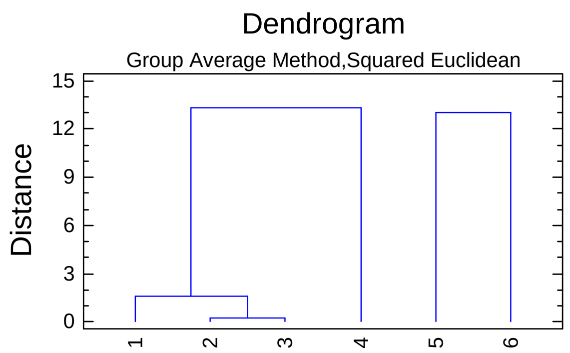
Seleccionamos las opciones dadas en la siguiente gráfica:



La representación gráfica de los datos es la siguiente:



El programa selecciona los cuatro primeros datos en un grupo y los dos últimos en el otro. El siguiente dendograma muestra el orden en que se han realizado los agrupamientos, comenzando por los puntos más cercanos. Primero ha agrupado los datos 2 y 3 en un cluster. Despues se ha incorporado a este cluster el dato más cercano que es el 1. Posteriormente se ha fomado un segundo cluster con los datoa 5 y 6. Por último se ha incorporado el dato 4 al primero de los clusters.



16.4 Análisis Factorial

Ejercicio 185 *Usando los datos de las variables* acel, cilindros, litroscien, peso, precio, y potencia *del fichero de datos* coches.sf.

1. *Realiza un análisis factorial para seleccionar un número de factores que parezca adecuado para representar las características de estos vehículos reteniendo al menos el 90% de la información contenida en los datos.*

2. *Realiza una rotación de estos factores para facilitar su interpretación.*

3. *Especifica los pesos de las variables primitivas en estos factores.*

4. *Da la ecuación expresión del primer factor en función de las variables primitivas.*

5. *Da los valores de estos factores para el primer vehículo.*

6. *Intenta dar un nombre a cada uno de estos factores, ya que pretendemos usarlos para describir los vehículos.*

1. La siguiente tabla muestra la salida de Statgrahics para un análisis factorial de los datos de las variables indicadas usando el procedimiento de las componentes principales, considerando seis factores (igual número que variables primitivas) y por tanto retienen entre los seis la información total sobre los datos.

Factor	Autovalor	% de varianza	% acumulado
1	3,46634	57,772	57,772
2	1,16572	19,429	77,201
3	0,87379 1	14,562	91,763
4	0,314423	5,240	97,003
5	0,121908	2,032	99,035
6	0,0578922	0,965	100,000

Utilizando tres factores conseguimos el 91.763% >90% de la información contenida en los datos:

2. Realizando una rotación varimax obtenemos los siguientes pesos de las variables primitivas en cada uno de los factores rotados.

	Factor 1	Factor 2	Factor 3
Acel	-0,119971	<u>0,986126</u>	0,0581521
Cilindros	<u>0,895844</u>	-0,0927902	-0,00118481
litroscien	<u>-0,896006</u>	0,121856	0,0632989
peso	<u>0,962437</u>	0,0664216	0,167614
precio	0,0468517	0,048697	<u>0,996285</u>
potencia	<u>0,862259</u>	-0,423756	0,0610435

3. Factor 1=-0,119971*acel + 0,895844*cilindros - 0,896006*litroscien + +0,962437*peso + 0,0468517*precio + 0,862259*potencia

4. El primer vehículo tendría los siguientes valores para cada una de las nuevas variables, los tres factores rotados: (-5.14793, 2.91549, -1.13788)

5. En la última tabla se han subrayado, para cada variable primitiva el valor mayor de sus pesos, para ver en que factor está mejor representada. De esta forma vemos que el primer factor, cuya expresión está dada anteriormente, tomará valores altos si son altos los valores de cilindros, peso, y potencia y un valor pequeño para litroscien, que representa el consumo de gasolina. dadas las características de las variables que intervienen con mayor peso en el factor 1, parece que un valor alto en el factor1 indicaría un buen diseño, así que podíamos dar este nombre a ese factor:Índice de *Calidad del Diseño.* La variable primitiva acel, que significa el tiempo en que el vehículo pasa de velocidad 0 a 60, es la que contribuye principalmente al segundo factor. También interviene moderadamente la potencia, aunque con coeficiente negativo. Tendrían un valor alto en este factor los vehículos que reaccionen rapídamente y tengan poca potencia. Son vehículos ágiles. Le llamo al factor 2 *Índice de Agilidad.* En el último factor influye casi exclusivamente el precio, ya que las otras variables tienen muy poco peso en este factor. Llamo a este factor 3 *Índice de precio.*

Unidad Temática VI

PROBLEMAS DE SERIES TEMPORALES

T. 17

Series temporales. Módelos clásicos

Ejercicio 186 *Calcular la media, la varianza, los dos primeros coeficientes de autocovarianza y los dos primeros coeficientes de autocorrelación de la serie cuyos valores son:*

$$-0.22, 0.27, -0.37,\ 0.15,\ 0.28,$$

$$0.15,\ 0.06,\ -0.34,\ 0.24,\ 0.02,\ 0.06$$

Calcular también los intervalos de confianza, al 95%, para estos dos coeficientes y contrasta la hipótesis de que cada uno de ellos sea nulo.

Media muestral:

$$\overline{x} = \frac{1}{N}\sum_{t=1}^{N} x_t,$$

$$\overline{x} = \frac{-0.22+0.27-0.37+0.15+0.28+0.15+0.06-0.34+0.24+0.02+0.06}{11} = \frac{0.3}{11} = 0.027273$$

Varianza muestral:

$$S_0^2 = \frac{1}{N}\sum_{t=1}^{N}(x_t - \overline{x}) =$$

$$\frac{(-0.22-0.027273)^2+(0.27-0.027273)^2+...+(0.02-0.027273)^2+(0.06-0.027273)^2.}{11} =$$

$$\frac{0.55318}{11} = 0.050289$$

Autocovarianza muestral de orden 1 :

$$S_1 = \frac{1}{11-1}\sum_{t=1}^{10}(x_t - \overline{x})(x_{t+1} - \overline{x})^2$$

$\sum_{t=1}^{10}(x_t-\overline{x})(x_{t+1}-\overline{x})^2=(-0.22-0.02727)(0.27-0.02727)+$
$+(0.27-0.02727)(-0.37-0.02727)+......+(0.24-0.02727)(0.02-0.02727)+$
$+(0.02-0.02727)(0.06-0.02727)=-0.23110;$

$$S_1=\frac{-0.231\,10}{10}=-0.023\,11$$

Autocovarianza muestral de orden 2 :

$$S_2=\frac{1}{11-2}\sum_{t=1}^{9}(x_t-\overline{x})(x_{t+m}-\overline{x})^2$$

$\sum_{t=1}^{9}(x_t-\overline{x})(x_{t+m}-\overline{x})^2=(-0.22-0.027273)(-0.37-0.027273)+$
$+(0.27-0.027273)(0.15-0.027273)+....+(0.24-0.27273)(0.06-0.27273)$
$=0.17858$

$$S_2=\frac{0.178\,58}{9}=0.019842$$

Coeficiente de Autocorrelacion de orden 1:

$$r_1=\frac{\sum_{t=1}^{10}(x_t-\overline{x})(x_{t+m}-\overline{x})^2}{\sum_{t=1}^{11}(x_t-\overline{x})^2}=\frac{-0.231\,10}{0.5318}=-0.434\,56$$

Coeficiente de Autocorrelación de orden 2 :

$$r_2=\frac{\sum_{t=1}^{9}(x_t-\overline{x})(x_{t+m}-\overline{x})^2}{\sum_{t=1}^{11}(x_t-\overline{x})^2}=\frac{0.001788}{0.5318}=3.\,362\,2\times10^{-3}=0.003362$$

Los intervalos de confianza para cada uno de estos intervalos son:
$\left(-\frac{1}{\sqrt{N}}F^{-1}_{N(0,1)}(1-\frac{\alpha}{2}),\ \frac{1}{\sqrt{N}}F^{-1}_{N(0,1)}(1-\frac{\alpha}{2})\right)=$
$=\left(-\frac{1}{\sqrt{11}}F^{-1}_{N(0,1)}(0.975),\ \frac{1}{\sqrt{11}}F^{-1}_{N(0,1)}(0.975)\right)=$
$=\left(-\frac{1}{\sqrt{11}}\times1.96,\ \frac{1}{\sqrt{11}}\times1.96\right)=(-0.59,0.59)\,.$

Como ambos coeficientes de autocorrelación quedan dentro de estos intervalos de confianza admitimos que ambos son nulos. No obstante, este resultado habría que ponerlo en entredicho a causa del pequeño tamaño muestral.

Ejercicio 187 *Las temperaturas medias registradas en una determinada localidad durante los meses de 4 años han sido las siguientes:*

MESES	*2000*	*2001*	*2002*	*2003*
Enero	*4*	*5*	*5*	*3*
Febrero	*10*	*9*	*11*	*12*
Marzo	*15*	*15*	*13*	*13*
Abril	*17*	*17*	*17*	*18*
Mayo	*18*	*19*	*18*	*19*
Junio	*21*	*20*	*22*	*23*
Julio	*27*	*27*	*27*	*27*
Agosto	*27*	*28*	*26*	*28*
Septiembre	*19*	*18*	*19*	*17*
Octubre	*12*	*13*	*11*	*10*
Noviembre	*9*	*9*	*8*	*8*
Diciembre	*5*	*6*	*6*	*6*

Calcular los coeficientes estacionales y predecir la temperatura media en Enero de 2004. Justificar la bondad de la predicción. Los datos están registrados en la variable cuatro *del fichero* ejemst.sf3*.*

Obtenemos la serie suavizada que corresponde a las medias móviles de orden 12 que se obtiene con la siguiente fórmula:

$$MM_t(12) = \frac{0.5x_{t-6} + \displaystyle\sum_{i=t-5}^{t+5} x_i + 0.5x_{t+6}}{12}$$

A continuación se obtienen los índices brutos de variación estacional, $IBVE$, dividiendo la serie primitiva por los valores correspondientes para la serie de medias móviles.

$$IBVE = \frac{x_t}{MM_t(12)}$$

Se calculan los índices (IVE) correspondientes a los términos que tienen el mismo número de orden dentro de cada período.

$$IVE_k = \frac{\sum_{i=1}^{12} IBVE_{12(i-1)+k}}{s-1}$$

donde $s = 4$

A continuación se normalizan los índices IVE_k dividiendo cada uno de ellos por la media de todos

$$IVEN_k = \frac{IVE_k}{\frac{\sum_{i=1}^{12} IVE_i}{12}}$$

Por último para desestacionalizar la serie se divide cada término de la serie por el correspondiente índice estacional normalizado.

Todo esto aparece recogido en la siguiente tabla:

Xt	MMt(4)	IBVE	IVE	IVEN	Serie desestac
4			0,27913959	0,27686911	14,4472599
10			0,68431267	0,67874658	14,7330393
15			0,87921091	0,87205955	17,2006602
17			1,11809412	1,10899972	15,3291293
18			1,2074367	1,1976156	15,0298643
21			1,48660925	1,4745174	14,2419479
27	15,375	1,75609756	1,75941761	1,74510678	15,4718326
27	15,375	1,75609756	1,75729909	1,74300549	15,4904847
19	15,3333333	1,23913043	1,21565263	1,20576471	15,7576349
12	15,3333333	0,7826087	0,78153535	0,77517846	15,4803063
9	15,375	0,58536585	0,56357895	0,55899489	16,1003261
5	15,375	0,32520325	0,36611967	0,36314171	13,7687296
5	15,3333333	0,32608696			18,0590749
9	15,375	0,58536585			13,2597354
15	15,375	0,97560976			17,2006602
17	15,375	1,10569106			15,3291293
19	15,4166667	1,23243243			15,8648568
20	15,4583333	1,29380054			13,5637599
27	15,5	1,74193548			15,4718326
28	15,58333334	1,7967914			16,0642064
18	15,5833333	1,15508021			14,9282857
13	15,5	0,83870968			16,7703318
9	15,4583333	0,58221024			16,1003261
6	15,5	0,38709677			16,5224755
5	15,5833333	0,32085561			18,0590749
11	15,5	0,70967742			16,2063432
13	15,4583333	0,84097035			14,9072389
17	15,4166667	1,1027027			15,3291293
18	15,2916667	1,17711172			15,0298643

22	15,25	1,44262295			14,9201359
27	15.1666667	1,78021978			15,4718326
26	15,125	1,71900826			14,9167631
19	15,1666667	1,25274725			15,7576349
11	15,2083333	0,72328767			14,1902807
8	15,2916667	0,52316076			14,3114009
6	15,5416667	0,38605898			16,5224755
3	15,75	0,19047619			10,835445
12	15,8333333	0,75789474			17,6796471
13	15,8333333	0,82105263			14,9072389
18	15,70833331	14588859			16,2308428
19	15,6666667	1,21276596			15,8648568
27	15,6666667	1,72340426			18,3110759
27					15,4718326
28					16,0642064
17					14,0989365
10					12,9002552
8					14,3114009
6					16,5224755

La recta de regresión de la serie desetacionalizada sobre t resulta:

$$y = -0.00344t + 15.52$$

Utilizando esta recta la temperatura estimada para enero de 2004 es:

$$\widehat{X_{49}} = \widehat{T_{49}} \times IVEN_{49} = (-0.00344 \times 49 + 15.52)\, 0.27690 = 4.25$$

Concluimos que para Enero de 2004 se espera una temperatura de 4,25. Una forma de decidir la fiabilidad de la predicción es comparar este valor con la desviación típica de los residuos. En este caso el valor es 1.25, que se considerá el error más probable.

Otra opción razonable sería considerar el modelo de media constante para la tendencia. En este caso la predicción para Enero de 2004 sería:

$$\widehat{X_{49}} = \widehat{T_{49}} \times IVEN_{49} = (15.44)\, 0.27690 = 4.28$$

El error cuadrático medio del ajuste es ahora 1.27.

Ejercicio 188 *Un laboratorio farmacológico presenta las siguientes cifras de ventas en las cuatro estaciones de cinco años:*

	VENTA EN MILLONES				
Primavera	*2.1*	*2.3*	*2.2*	*2.5*	*2.6*
Verano	*3.2*	*3.1*	*3.6*	*3.7*	*3.7*
Otoño	*2.6*	*2.9*	*4.4*	*4.5*	*4.9*
Invierno	*1.4*	*1.6*	*1.7*	*1.8*	*2.1*

Obtener los coeficientes estacionales y hacer una predicción de las ventas en la próxima primavera.

Obtenemos la serie suavizada que corresponde a las medias móviles de orden 4 que se obtiene con la siguiente fórmula:

$$MM_t(4) = \frac{0.5x_{t-2} + x_{t-1} + x_t + x_{t+1} + 0.5x_{t+2}}{4}$$

A continuación se obtienen los índices brutos de variación estacional, $IBVE$, dividiendo la serie primitiva por los valores correspondientes para la serie de medias móviles.

$$IBVE = \frac{x_t}{MM_t(4)}$$

Se calculan los índices (IVE) correspondientes a los términos que tienen el mismo número de orden dentro de cada período.

$$IVE_k = \frac{\sum_{i=1}^{4} IBVE_{4(i-1)+k}}{s-1}$$

donde $s = 5$

A continuación se normalizan los índices IVE_k dividiendo cada uno de ellos por la media de todos

$$IVEN_k = \frac{IVE_k}{\frac{\sum_{i=1}^{4} IVE_i}{4}}$$

Por último para desestacionalizar la serie se divide cada término de la serie por el correspondiente índice estacional normalizado.

Todo esto aparece recogido en la siguiente tabla:

Xt	MMt(4)	IBVE	IVE	IVEN	Serie desestac
2,1			0,84548687	0,86101173	2,438991154
3,2			1,19868158	1,22069181	2,621464289
2,6	2,35	1,10638298	1,29472295	1,31849671	1,97194273
1,4	2,3625	0,59259259	0,58898478	0,59979974	2,6712760268
2,3	2,3875	0,96335079			2,33411236
3,1	2,45	1,26530612			2,53954353
2,9	2,4625	1,17766497			2,19947458
1,6	2,5125	0,63681592			2,667556992
2,2	2,7625	0,79638009			2,55513359
3,6	2,9625	1,21518987			2,949147325
4,4	3,0125	1,46058091			3,337133852
1,7	3,0625	0,55510204			2,834279304
2,5	3,0875	0,8097166			2,903560898
3,7	3,1125	1,18875502			3,031068084
4,5	3,1375	1,43426295			3,412977803
1,8	3,15	0,57142857			3,001001616
2,6	3,2	0,8125			3,019703334
3,7	3,2875	1,12547529			3,031068084
4,9					3,716353607
2,1					3,501168552

La recta de regresión que se ajusta a la serie desestacionalizada resulta:

$$y = 0.061t + 2.1955$$

Utilizando esta recta la predicción de las ventas para la próxima primavera será:

$$\widehat{X_{21}} = \widehat{T_{21}} \times IVEN_{21} = (0.061 \times 21 + 2.1955)\, 0.861 = 2.993$$

Concluimos que para la próxima primavera se esperan unas ventas de 2,993 millones

Ejercicio 189 *La tabla adjunta contiene el número de nacimientos habidos en España (en miles) entre los años 1967 a 1971 inclusive, agrupados por cuatrimestres:*

	1967	*1968*	*1969*	*1970*	*1971*
1^{er} *cuatrim.*	*57*	*59*	*60*	*62*	*67*
2^{o} *cuatrim.*	*82*	*96*	*107*	*118*	*129*
3^{o} *cuatrim.*	*80*	*88*	*91*	*92*	*96*

1. *Desestacionalizar la serie usando el método de la razón a la media móvil.*

2. *Aplicando un esquema adecuado, estudiar la tendencia.de la serie.*

3. *Hacer una predicción del número de nacimentos para el segundo cuatrimestre de 1972.*

Si observamos los datos en la tabla, vemos que dentro de cada año natural el valor más alto corresponde al segundo cuatrimestre, mientras que el más bajo corresponde al primer cuatrimestre. El valor del tercer cuatrimestre crece muy lentamente y tiende, con el transcurso de los años, a "equidistar" de los cuatrimestres que le flanquean. Si por el contrario nos fijamos en las filas, apreciamos que los valores son crecientes según transcurren los años.

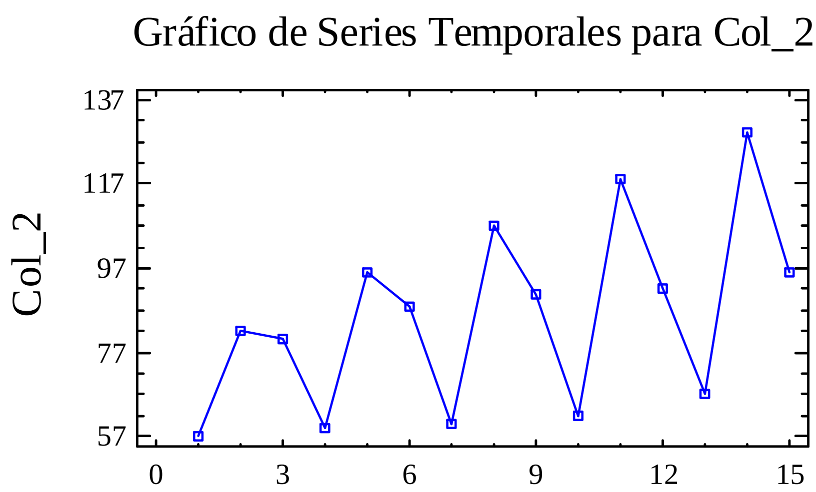


Figura 17.1:

La visión de los valores de la serie en el gráfico de la figura 17.1, nos inclina a pensar en aplicar el siguiente procedimiento:

1^o Desestacionalizaremos la serie.

La salida de la hoja de cálculo Excel es:

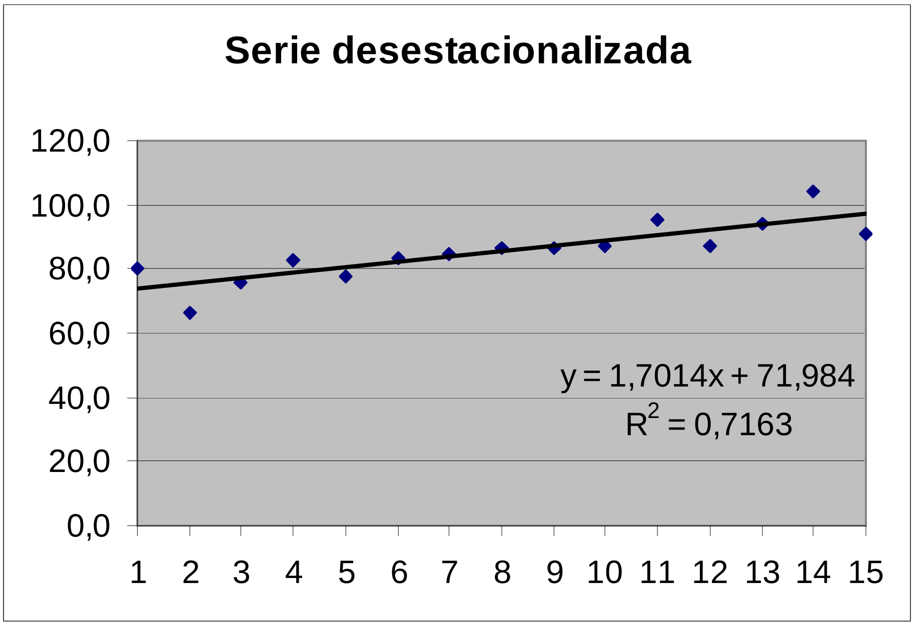
Xt	$Tt = MMt(3)$	$IBVEt$	IVE	$IVEN$	$S.Desest$
57			0,7108	0,710761	80,2
82	73,0	1,1233	1,2359	1,235766	66,4
80	1,0860	73,7	1,0536	1,053473	75,9
59	78,3	0,7532	0,7108	0,710761	83,0
96	81,0	1,1852	1,2359	1,235766	77,7
88	81,3	1,0820	1,0536	1,053473	83,5
60	85,0	0,7059	0,7108	0,710761	84,4
107	86,0	1,2442	1,2359	1,235766	86,6
91	86,7	1,0500	1,0536	1,053473	86,4
62	90,3	0,6863	0,7108	0,710761	87,2
118	90,7	1,3015	1,2359	1,235766	95,5
92	92,3	0,9964	1,0536	1,053473	87,3
67	96,0	0,6979	0,7108	0,710761	94,3
129	97,3	1,3253	1,2359	1,235766	104,4
96				1,053473	91,1

ECM=18.81

Obtenidos los Índices de Variación Estacional Normalizados, podemos obtener los valores de la serie, ya desestacionalizados.

2^o Estimaremos la tendencia, que parece ser lineal con correlación positiva.

A la vista de la siguiente gráfica, se confirma la sospecha inicial de una tendencia lineal, por lo que procedemos a calcular la recta de regresión.



La recta de regresión ajustada a la serie desestacionalizada resulta: $y = 1,7014x + 71,984$

La estimación para el segundo cuatrimestre de 1972 es:
$\widehat{X_{17}} = T_{17} \times IVEN_2 = (1.7014 * 17 + 71.984) \times 1.235766 = 124.70$

Están previstos unos 124.700 nacimientos para el segundo cuatrimestre de 1972.

El error cuadrático medio de este ajuste es 4.76.

Ejercicio 190 *La tabla adjunta, idéntica a la del ejercicio 189, contiene el número de nacimientos habidos en España (en miles) entre los años 1967 a 1971 inclusive, agrupados por cuatrimestres:*

	1967	*1968*	*1969*	*1970*	*1971*
1^{er} *cuatrim.*	*57*	*59*	*60*	*62*	*67*
2^{o} *cuatrim.*	*82*	*96*	*107*	*118*	*129*
3^{o} *cuatrim.*	*80*	*88*	*91*	*92*	*96*

Aplica el método de Holt-Winters para desestacionalizar la serie y emplea el modelo estimado parahacer una predicción del número de nacimentos para el segundo cuatrimestre de 1972.Tomad para los tres parámetros α, β, γ el valor 0.1.

Tomamos como valores de partida los valores de a_0 y b_0 del ejercicio 189. También tomamos como valores de partida los indices estacionales obtenidos en este mismo ejercicio. Con estos valores de partida e utilizando las expresiones propias de este método:

$$a_t = \alpha \frac{X_t}{E_{t-L}} + (1-\alpha)(a_{t-1}+b_{t-1}) \quad 0<\alpha<1$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \quad 0<\beta<1$$

$$E_t = \gamma \frac{X_t}{a_t} + (1-\gamma)E_{t-L} \quad 0<\gamma<1$$
$$\widehat{X}_{t+m}/t = (a_t + b_t m)\, E_{t+m-L}$$

Las predicciones de pasado se realizan con la fórmula:

$$\widehat{X}_{t/t-1} = (a_{t-1}+b_{t-1})\, E_{t-L}$$

se obtiene la tabla siguiente:

X_t	a_t	b_t	E$_t$	$\widehat{X_t}$
			0.710761	
			1.235766	
	71.984	1.7014	1.053473	
57	74.33643339	1.766503339	0.7396849	52.37270859
82	75.12820351	1.669030017	1.2121894	94.04542171
80	76.71143996	1.660450661	1.0481257	80.90381199
59	78.51107096	1.674368694	0.740865047	57.97050408
96	80.08645017	1.664469746	1.210840925	97.19993998
88	81.97176746	1.6865545	1.050667165	85.68524017
60	83.3911307	1.659835374	0.73872864	61.97952666
107	85.38270326	1.693009093	1.215074954	102.9831904
91	87.02930452	1.68836831	1.05016293	91.48759184
62	88.2387032	1.640471347	0.735119728	65.53828582
118	90.60259215	1.712813107	1.223806558	109.2099339
92	91.84441011	1.665713593	1.045316043	96.94621643
67	93.27327294	1.642028516	0.733439694	68.74113672
129	95.96465289	1.746963659	1.235850397	116.1579684
96	97.12428027	1.688230031	1.03962687	102.1395203
				72.4730173
				124.2038798

RECM=$\sqrt{\frac{\sum_{t=1}^{15}\left(X_{tt}-\widehat{X_t}\right)^2}{15}} = 5.861527543$

Ejercicio 191 *La tabla siguiente muestra las produciones mensuales medias de maiz, en millones de toneladas para los años 1948-1958.*

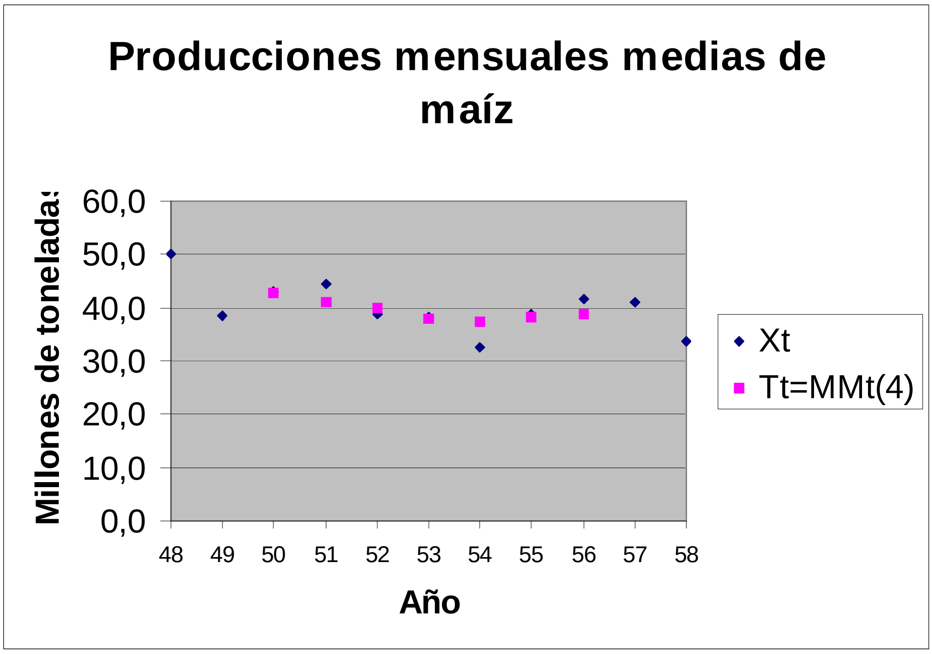
48	*49*	*50*	*51*	*52*	*53*	*54*	*55*	*56*	*57*	*58*
50'0	*38'5*	*43'0*	*44'5*	*38'9*	*38'1*	*32'6*	*38'7*	*41'7*	*41'1*	*33'8*

1. *Construir las medias móviles de cuatro años para obtener la tendencia.*

2. *Representar gráficamente los datos originales y los de las medias móviles en un mismo gráfico.*

3. *Proponer modelos que parezcan adecuados para representar la tendencia de la serie.*

Año	Xt	Tt=MMt(4)
48	50,0	
49	38,5	
50	43,0	42,6
51	44,5	41,2
52	38,9	39,8
53	38,1	37,8
54	32,6	37,4
55	38,7	38,2
56	41,7	38,7
57	41,1	
58	33,8	



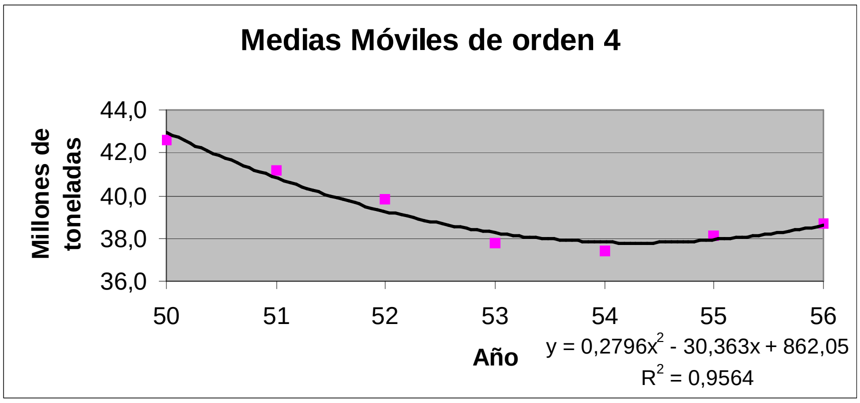
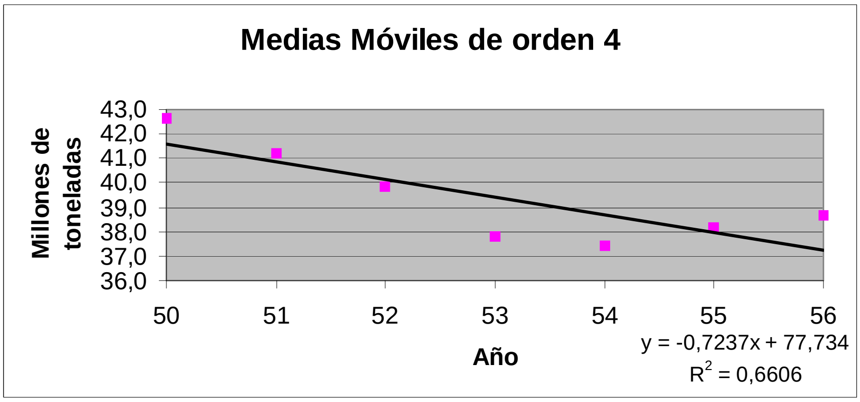


Figura 17.2:

Los datos relativos a la tendencia sugieren que, en principio, un polinomio de orden 2 sería más adecuado que una recta.



Para el ajuste lineal se obtiene un R-cuadrado igual a 0.66, y lo que es peor, una imagen pobre de adaptación de la recta a la nube de puntos.

La parábola se ajusta perfectamente a la nube de puntos. El R-cuadrado es superior a 0.95. Es claramente satifactoria.

Ejercicio 192 *Hallar un modelo de suavizado por el método de Holt para la serie del ejercicio 191.*

Tomando $\alpha = 0.2$, y $S_0 = 50$ se obtiene el siguiente suavizado

X_t	S_t
	50
50	50
38.5	47.7
43	46.76
44.5	46.308
38.9	44.8264
38.1	43.48112
32.6	41.304896
38.7	40.7839168
41.7	40.96713344
41.1	40.99370675
33.8	39.5549654

Ejercicio 193 *Hallar un modelo de suavizado por el método de Brown para la serie del ejercicio 191 .*

Tomando $\alpha = 0.2$, y $a_0 = 50$, $b_0 = 0$ obtiene el siguiente suavizado

X_t	S_t	S'_t	a_t	b_t	$\widehat{X_t}$
	50	50	50	0	
50	50	50	50	0	50
38.5	47.7	49.54	45.86	-0.46	50
43	46.76	48.984	44.536	-0.556	45.4
44.5	46.308	48.4488	44.1672	-0.5352	43.98
38.9	44.8264	47.72432	41.92848	-0.72448	43.632
38.1	43.48112	46.87568	40.0865	-0.84864	41.204
32.6	41.304896	45.7615232	36.848269	-1.114157	39.23792
38.7	40.7839168	44.7660019	36.801832	-0.995521	35.73411
41.7	40.96713344	44.0062282	37.928038	-0.7597737	35.80631
41.1	40.99370675	43.4037239	38.583690	-0.6025043	37.16827
33.8	39.5549654	42.6339722	36.475959	-0.7697517	37.98119

Ejercicio 194 *¿Cuál de los modelos, tratados en los ejercicios 191 y siguientes, sobre las producciones anuales de maiz está mejor ajustado?.*

Usando como criterio de comparación la desviación típica de los residuos, el mejor ajuste es el de la recta. Los valores son los siguientes.

$RECM_{\text{recta}} = 3.87$

$RECM_{\text{parábola}} = 4.06$

$RECM_{\text{Holt}} = 5.044448573$

$RECM_{\text{Brown}} = 5.142001215$

Ejercicio 195 *Se realizó un estudio de seguimiento de la cantidad de insectos de una cierta especies recontados en un espacio natural protegido obteniéndose los siguientes datos.*

	Pri.	*Ver.*	*Oto.*	*Inv.*
1999	*203*	*424*	*82*	*506*
2000	*301*	*501*	*163*	*607*
2001	*342*	*588*	*184*	*669*

1. *Suponiendo que la serie temporal se ajusta al esquema aditivo, desestacionalizar los valores de la serie.*

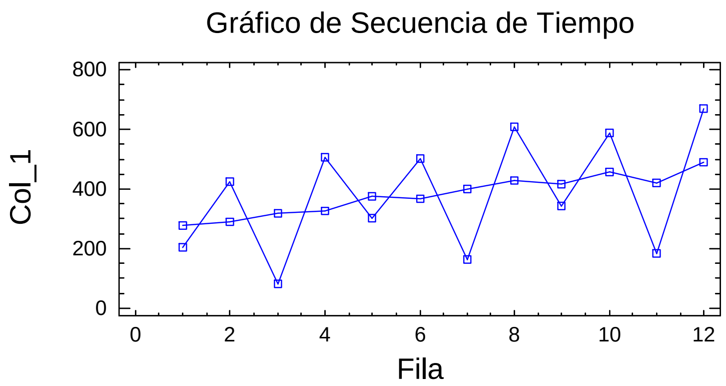
2. *Modelar la serie y predecir el número de insectos que habrá en el verano de 2002.*

El esquema del modelo aditivo es: $X_t = T_t + E_t + \varepsilon_t$. Para ajustar este modelo seguimos un procedimiento similar al usado cuando el método de la razón a la media movil: $X_t = T_t \times E_t + \varepsilon_t = MM_t(l) \times E_t + \varepsilon_t$

X_t	$MM_t(4)$	IBVE	IVE	IVEN	Ser. Desest.
203			-71.125	-73.66625	276.66625
424			135.3125	132.77125	291.22875
82	316	-234	-234.5625	-237.10375	319.10375
506	337.875	168.125	180.5	177.95875	328.04125
301	357.625	-56.625	-71.125	-73.66625	374.66625
501	380.375	120.625	135.3125	132.77125	368.22875
163	398.125	-235.125	-234.5625	-237.10375	400.10375
607	414.125	192.875	180.5	177.95875	428.98125
342	427.625	-85.625	-71.125	-73.66625	415.66625
588	438	150	135.3125	132.77125	455.22875
184			-234.5625	-237.10375	421.10375
669			180.5	177.95875	490.98125

media de los IVE = 2.54125

La siguiente figura superpone la serie primitiva y la serie desestacionalizada.



Ajustando una tendencia lineal a la serie desetacionalizada se obtiene $\mathrm{T}_t = 17.874T + 264.648$ con un coeficiente de correlación $r = 0.9669$

El modelo de la serie es:

$X_t = (17.874t + 264.648) + E_{\mathrm{mod}(t,4)}$

La previsión para el verano de 2002 será: $X_{14} = (17.874t + 264.648) + E_2 = (17.874 \times 14 + 264.648) + 132.77 = 647.65$. Habrá, más o menos, 468 insectos en el verano de 2002.

Ejercicio 196 *Consideraremos la serie temporal de la siguiente tabla, tomada de los datos del Instituto Nacional de estadística dentro de la sección de* Hostelería y turismo *de la página* http://www.ine.es/inebase/ *La tabla registra el número de entradas de personas que visitan nuestro pais (datos mensuales en miles de personas). Los datos están en la variable* totalvisitantes *del fichero* ejemst.sf3

periodo	*visitantes*	*periodo*	*visitantes*	*periodo*	*visitantes*
1999M02	*3728.7*	*2000M02*	*3920.1*	*2001M02*	*4091.7*
1999M03	*4613.3*	*2000M03*	*4804.1*	*2001M03*	*4897.7*
1999M04	*5627.4*	*2000M04*	*6533.2*	*2001M04*	*6588*
1999M05	*6569.8*	*2000M05*	*6185.5*	*2001M05*	*6453.4*
1999M06	*6270.6*	*2000M06*	*6723.5*	*2001M06*	*6972.1*
1999M07	*9500.9*	*2000M07*	*9561*	*2001M07*	*9641.5*
1999M08	*10399.5*	*2000M08*	*10325.2*	*2001M08*	*10761.3*
1999M09	*6906.9*	*2000M09*	*7688.8*	*2001M09*	*7492.8*
1999M10	*6319.1*	*2000M10*	*6230.8*	*2001M10*	*6002*
1999M11	*4227.7*	*2000M11*	*4312.6*	*2001M11*	*4209.2*
1999M12	*4300.7*	*2000M12*	*4552.6*	*2001M12*	*4666.6*
2000M01	*3624.4*	*2001M01*	*3901.9*	*2002M01*	*3925.8*

periodo	*visitantes*	*periodo*	*visitantes*
2002M02	*4424.8*	*2003M02*	*4423.8*
2002M03	*5785*	*2003M03*	*5545.7*
2002M04	*6039.1*	*2003M04*	*6712.6*
2002M05	*6789.4*	*2003M05*	*7378.7*
2002M06	*7131*	*2003M06*	*7510.3*
2002M07	*9869.8*	*2003M07*	*10117*
2002M08	*12199.3*	*2003M08*	*11847.4*
2002M09	*7629.4*	*2003M09*	*7652.8*
2002M10	*6528.7*	*2003M10*	*6791.2*
2002M11	*4720.1*	*2003M11*	*4907.7*
2002M12	*4982*	*2003M12*	*5358.4*
2003M01	*4279.5*	*2004M01*	*4673.9*

Realizar el estudio de la serie usando un modelo clásico

La primera fase del estudio de la serie es su representación gráfica que aparece en la figura 17.3

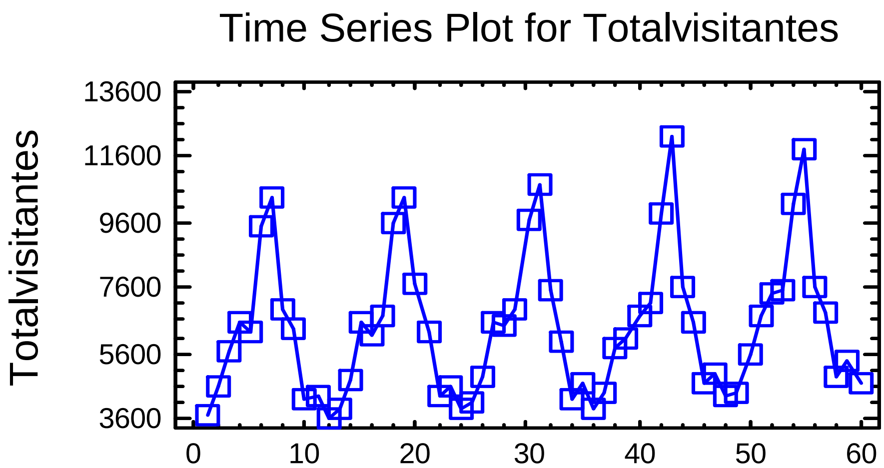


Figura 17.3:

Observamos una cierta pauta que se repite cada 12 meses. Por ejemplo, los valores más altos se dan en Agosto de cada año y los más bajos en Enero. Además se detecta un leve aumento (a largo plazo) de los valores de la serie correspondientes a un mismo mes. Para describir ambos aspectos de la evolución de esta serie a traves del tiempo le aplicaremos un modelo de descomposición aditivo, es decir, cada valor de la serie será suma de una componente de tendencia, una componente estacional que depende del mes y una componente residual que ha de ser puramente aleatoria:

$$X_t = T_t + E_t + \varepsilon_t$$

Comenzamos adoptando un modelo para representar la componente estacional E_t. Existen distintos métodos para definir esta componente Nos decidimos por un modelo de estacionalidad constante. Realizando este cálculo con Statgraphics, resulta la función periódica representada grafícamente en la figura 17.4.

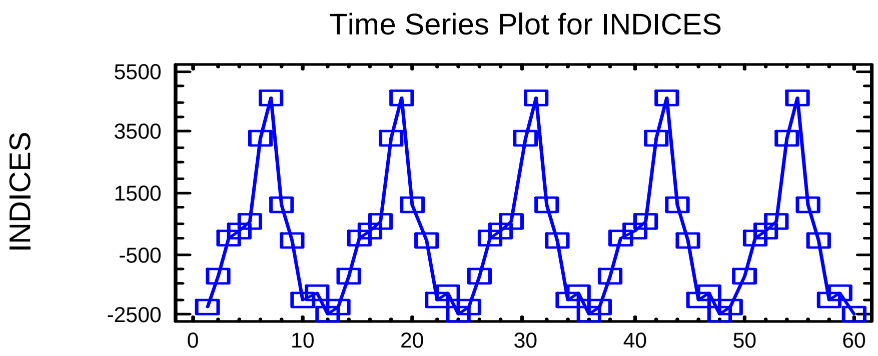
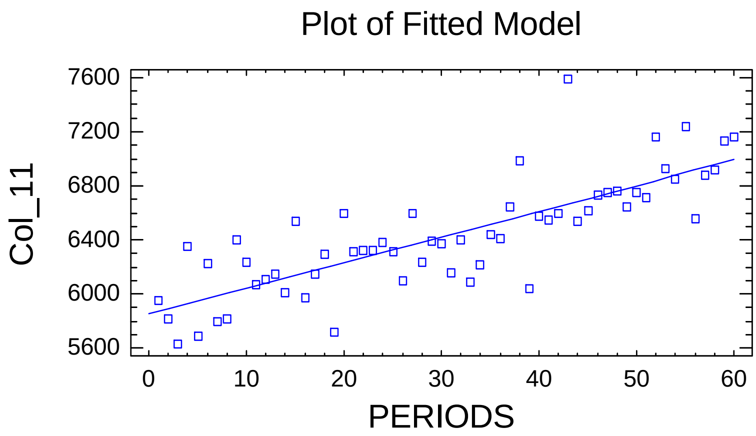


Figura 17.4:

Para representar la tendencia vamos elegir en este caso un modelo lineal. Para calcularla se emplea la recta de regresión ajustada a los puntos de la serie resultante de eliminar en la serie la componente estacional: $X_t - E_t$.

La gráfica de la figura siguiente



representa los valores de la serie $X_t - E_t$ y la recta de regresión ajustada a los puntos $(t, X_t - E_t)$, que nos va servir para representar la tendencia. El cálculo realizado con Statgraphics nos da el modelo lineal:

$$T_t = 5850.73 + 19.0079t$$

que está representado en la citada figura, donde los valores de t, número de orden del mes correspondiente dentro de la serie, están representados en el eje horizontal.

Para obtener la componente irregular basta calcular $\varepsilon_t = X_t - (T_t + E_t)$. Viene representada en la siguiente gráfica de la figura 17.5.

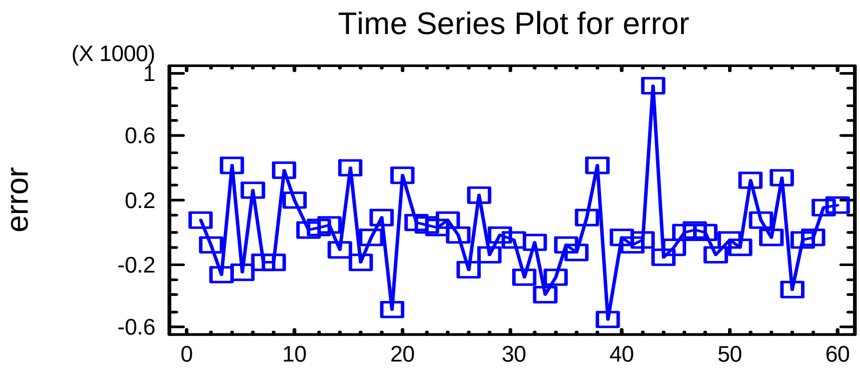


Figura 17.5:

Ahora tendríamos que decidir si este residual es puramente aleatorio o por el contrario todavía contiene alguna información. Realizando algunos tests de hipótesis se aceptan todos salvo uno. Parece que el modelo, si bien no es del todo adecuado, no resulta tampoco demasiado malo. Si nos decidimos por aceptar este modelo para describir la serie, podemos usarlo para hacer previsiones sobre el número de visitantes en el año próximo. Por ejemplo la previsión para el mes de Marzo de 2004 se calcularía de la forma siguiente:

$X_{62} = T_t + E_t = (5850.73 + 19.0079 \times 62) + (-1201.52) = 5827.7$

En la gráfica de la figura 17.6 están representados los valores de la serie primitiva, el valor calculado con el modelo para los cinco años de los datos. Se observa que el modelo actual describe la serie bastante bien. Tambien estan representados los valores previstos para el siguiente año con un intervalo de confianza al 95%. Para calcular este intervalo de confianza se usa como referencia los errores cometidos en el pasado empleando su desviación típica. En este caso el valor de la desviación típica es 246.108, valor que nos da una orientación sobre la magnitud de los errores esperados.

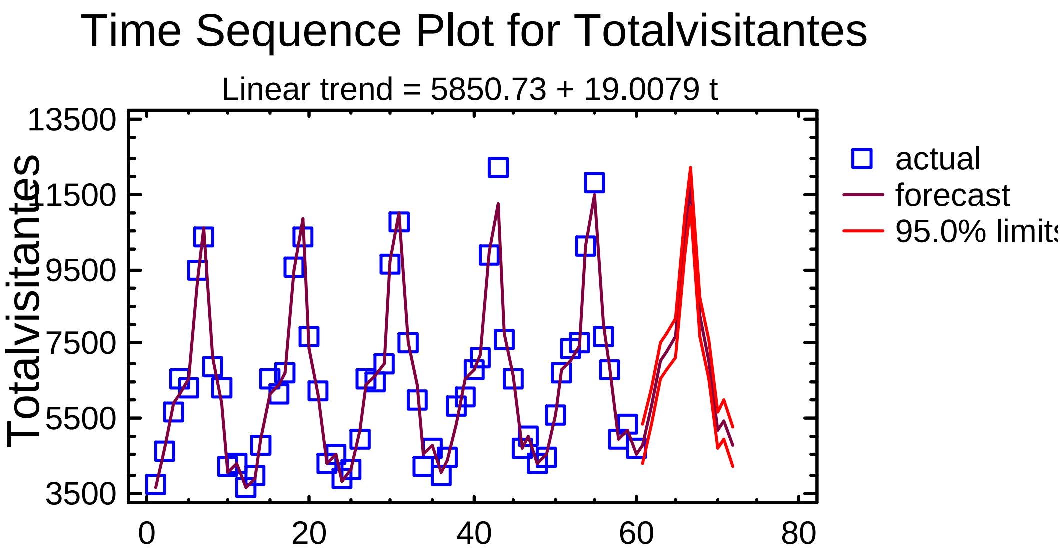


Figura 17.6:

T. 18

Series temporales. Módelos ARIMA

Ejercicio 197 *Las temperaturas medias registradas en una determinada localidad durante los meses de 4 años han sido las siguientes:*

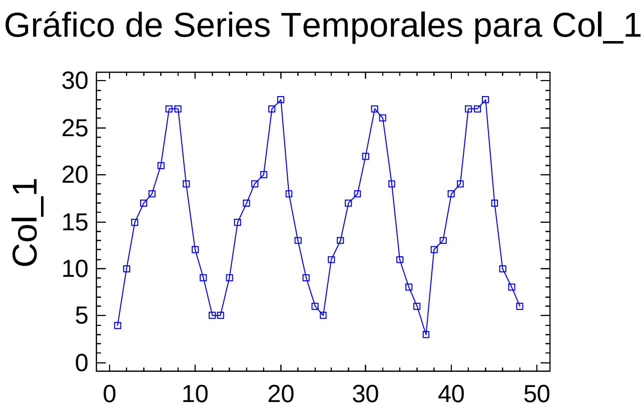
MESES	*2000*	*2001*	*2002*	*2003*
Enero	*4*	*5*	*5*	*3*
Febrero	*10*	*9*	*11*	*12*
Marzo	*15*	*15*	*13*	*13*
Abril	*17*	*17*	*17*	*18*
Mayo	*18*	*19*	*18*	*19*
Junio	*21*	*20*	*22*	*23*
Julio	*27*	*27*	*27*	*27*
Agosto	*27*	*28*	*26*	*28*
Septiembre	*19*	*18*	*19*	*17*
Octubre	*12*	*13*	*11*	*10*
Noviembre	*9*	*9*	*8*	*8*
Diciembre	*5*	*6*	*6*	*6*

Los datos están en la variable cuatro *del fichero* ejemst.sf3.

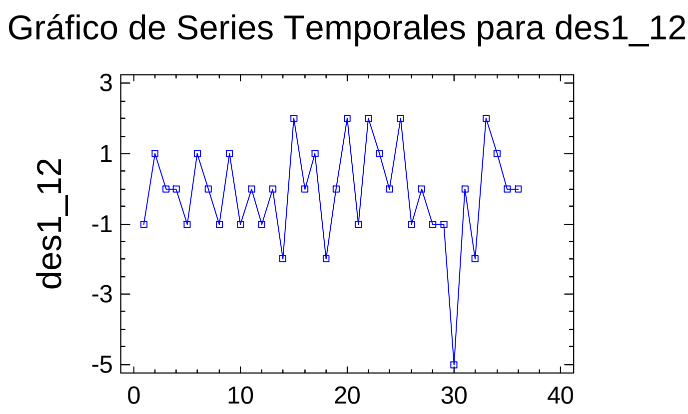
1. *Realiza una diferenciación estacional de periodo 12, para conseguir una serie desestacionalizada.*
2. *Cálcula los dos primeros coeficientes de autocorrelación total y parcial.*
3. *Por medio de algún programa estadístico, haz la representacion gráfica de la funciones de autocorrelación total y parcial.*
4. *¿Son compatibles estos correlogramas con la identificación del modelo AR(1) para la serie desestacionalizada?*

5. *Predecir la temperatura media en Enero de 2004. Justificar la bondad de la predicción.*

1. La serie dada, X_t se representa de la siguiente forma:



La serie $Y_t = X_t - X_{t+12}$ está representada en la figura siguiente, habiendo desaparecido la componente estacional.



2. Calculando los coeficientes de autocorrelación total por medio de la expresión $r_m = \frac{\sum_{t=1}^{N-m}\left(Y_t-\overline{Y}\right)\left(Y_{t+m}-\overline{Y}\right)^2}{\sum_{t=1}^{N}\left(Y_t-\overline{Y}\right)^2}$, se obtiene $r_1 = -0.14458$ y $r_2 = 0.2729$.

   Los coeficientes de autocorrelación parcial pueden obtenerse realizando un ajuste por regresión: Las variables que intervienen en la regresión son:

   1) $Z_t = Y_t - \overline{Y_t}$, siendo Y_t la serie obtenida previamente realizando un retardo de orden 12, $Y_t = B_{12}X_t = (X_t - X_{t+12})$,

   2) $DZ_t = Z_{t-1}$, obtenida de la anterior suprimiendo el primer elemento.

   3) $D^2Z_t = Z_{t-2}$, obtenida de Z_{t-1} suprimiendo el primer elemento.

   Las rectas de regresión ajustadas son

   $Z_t = \alpha_{11}Z_{t-1}$

   $Z_t = \alpha_{21}Z_{t-1} + \alpha_{22}Z_{t-2}$

   La primera resulta: $Z_t = 0.127Z_{t-1}$

   La segunda es: $Z_t = -0.0894Z_{t-1} + 0.2654$

   Por tanto los valores estimados por este procedimiento son: $\widehat{\phi_{11}} = -0.127$ $\widehat{\phi_{22}} = 0.2654$

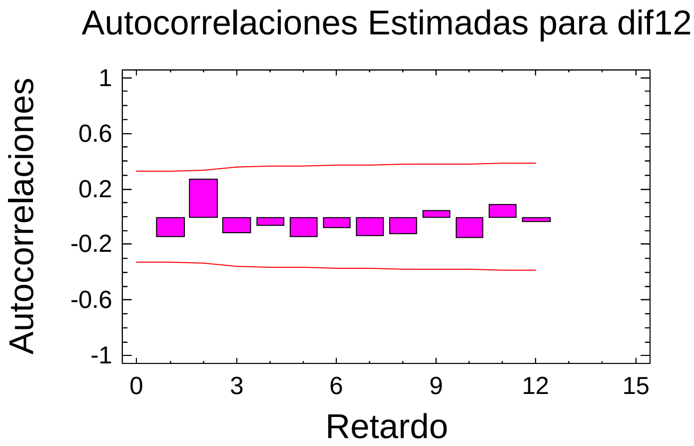
   Los valores obtenidos con Statgraphics 5.1 para la estimación de los dos primeros coeficientes de autocorrelación total son: $r_1 = -0.14458$, $r_2 = 0.2729$, y para los dos primeros coeficientes de autocorrelación parcial $\widehat{\phi_{11}} = -0.14458$ $\widehat{\phi_{22}} = 0.2574$. Estos valores están obtenidos por el procedimiente de Yule-Walker que nos da: $\widehat{\phi_{11}} = r_1 = -0.144582$ y $\widehat{\phi_{22}} = \phi_2 = 0.2574$, obtenido resolviendo el siguiente sistema:

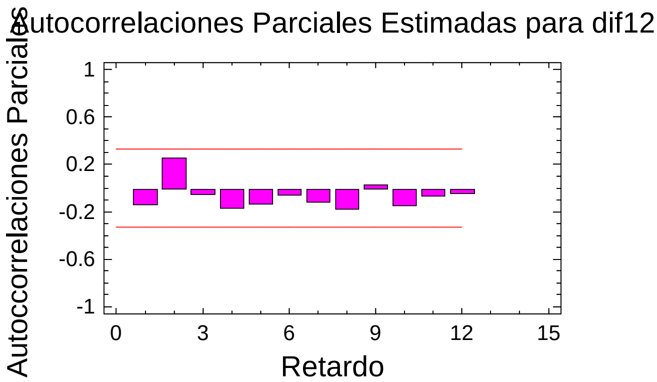
   $$\begin{pmatrix} 1 & -0.14458 \\ -0.14458 & 1 \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} -1.4458 \\ 0.2729 \end{pmatrix};$$

   $$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} 1 & -0.14458 \\ -0.14458 & 1 \end{pmatrix}^{-1}\begin{pmatrix} -0.14458 \\ 0.2729 \end{pmatrix} = \begin{pmatrix} -0.1074 \\ 0.2574 \end{pmatrix}$$

   Las diferencias entre los distintos procedimientos de estimación están justificadas, ya que los estimadores no son idénticos.

3. Los autocorrelogramas total y parcial son los siguientes





4. No es compatible con un proceso $AR(1)$,ya que estos procesos han de tener un primer coeficiente de autocorrelación parcial no nulo. Más bien sería compatible con un proceso de media constante ya que todos los coeficientes de autocorrelación pueden ser considerados nulos, pues quedan dentro de los intervalos de confianza.. El valor de la media no resulta significativo, así que tomamos un modelo de ruido blanco para la serie diferenciada.

5. Si aceptamos este modelo como plausible, $Y_t = \varepsilon_t$ la serie primitiva podría modelarse como $X_{t+12} = X_t + \varepsilon_t$. Por tanto la previsión para

Enero de 2004 sería $X_{49} = X_{37} = 3$. El error estimado en este modelo resulta 1.41, algo mayor que el del modelo usado en el ejercicio 197.

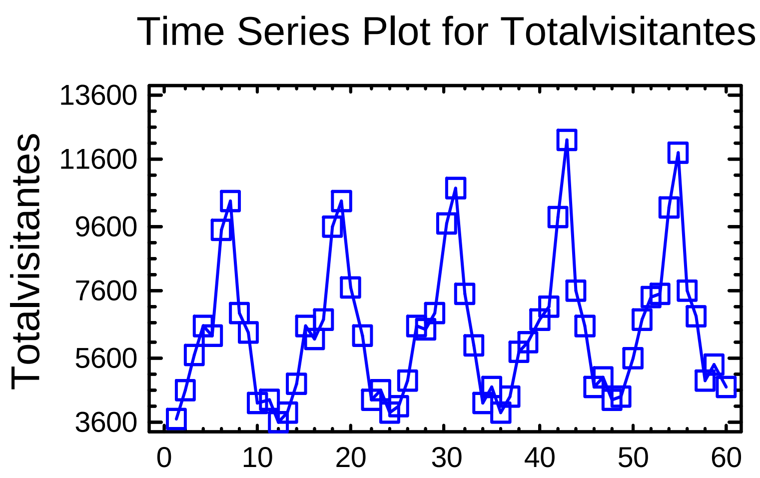
Ejercicio 198 *Consideraremos la serie temporal de la siguiente tabla, tomada de los datos del Instituto Nacional de estadística dentro de la sección de* Hostelería y turismo *de la página* http://www.ine.es/inebase/ *La tabla registra el número de entradas de personas que visitan nuestro pais (datos mensuales en miles de personas). Los datos están en la variable* totalvisitantes *del fichero* ejemst.sf3.

periodo	*visitantes*	*periodo*	*visitantes*	*periodo*	*visitantes*
1999M02	*3728.7*	*2000M02*	*3920.1*	*2001M02*	*4091.7*
1999M03	*4613.3*	*2000M03*	*4804.1*	*2001M03*	*4897.7*
1999M04	*5627.4*	*2000M04*	*6533.2*	*2001M04*	*6588*
1999M05	*6569.8*	*2000M05*	*6185.5*	*2001M05*	*6453.4*
1999M06	*6270.6*	*2000M06*	*6723.5*	*2001M06*	*6972.1*
1999M07	*9500.9*	*2000M07*	*9561*	*2001M07*	*9641.5*
1999M08	*10399.5*	*2000M08*	*10325.2*	*2001M08*	*10761.3*
1999M09	*6906.9*	*2000M09*	*7688.8*	*2001M09*	*7492.8*
1999M10	*6319.1*	*2000M10*	*6230.8*	*2001M10*	*6002*
1999M11	*4227.7*	*2000M11*	*4312.6*	*2001M11*	*4209.2*
1999M12	*4300.7*	*2000M12*	*4552.6*	*2001M12*	*4666.6*
2000M01	*3624.4*	*2001M01*	*3901.9*	*2002M01*	*3925.8*

periodo	*visitantes*	*periodo*	*visitantes*
2002M02	*4424.8*	*2003M02*	*4423.8*
2002M03	*5785*	*2003M03*	*5545.7*
2002M04	*6039.1*	*2003M04*	*6712.6*
2002M05	*6789.4*	*2003M05*	*7378.7*
2002M06	*7131*	*2003M06*	*7510.3*
2002M07	*9869.8*	*2003M07*	*10117*
2002M08	*12199.3*	*2003M08*	*11847.4*
2002M09	*7629.4*	*2003M09*	*7652.8*
2002M10	*6528.7*	*2003M10*	*6791.2*
2002M11	*4720.1*	*2003M11*	*4907.7*
2002M12	*4982*	*2003M12*	*5358.4*
2003M01	*4279.5*	*2004M01*	*4673.9*

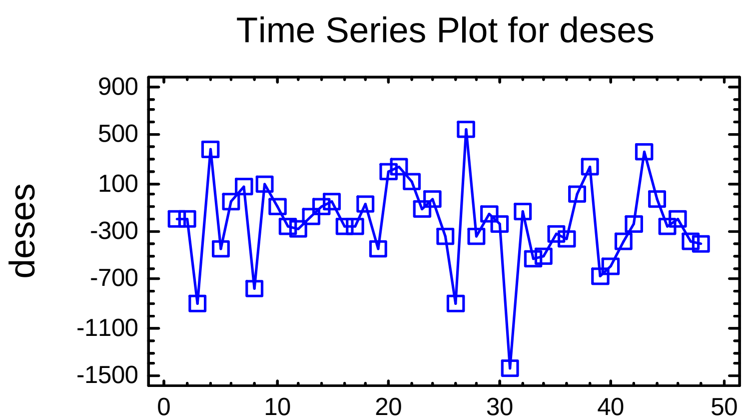
Realizar el estudio de la serie usando un modelo ARIMA. Esta serie es la misma que la del ejercicio 196.

La representación gráfica es la siguiente:



La serie del número de entradas de viajeros que tratamos de estudiar no es estacionaria, ya que la afluencia media de viajeros no es constante. Por lo pronto, se observa fácilmente que el número medio de viajeros varía según el mes del año.

Para eliminar la componente de estacionalidad se construye la serie $Y_t = X_t - X_{t-12}$, que está representada en la figura **??**. La operación realizada es una diferenciación estacional de periodo 12.



Esta serie parece ser de tendencia constante, es decir que admitimos que no hay excesivas variaciones en media ni en varianza y admitimos que se ajusta a un modelo estacionario. Tratemos de identificar ahora el modelo ARMA más adecuado. Para encontrar un modelo aceptable se estudian los primeros coeficientes de autocorrelación En este caso los primeros coeficientes de autocorrelación pueden considerarse nulos, por ello tomamos $p = 0$ y $q = 0$ y por tanto representamos la serie diferenciada de la gráfica por medio de un modelo de media constante. El modelo seleccionado para la serie es $Y_t = c + \varepsilon_t$. Una estimación de c es la media muestral.de Y_t , 225.635. Por tanto el modelo que vamos a emplear en las predicciones para la serie de entrada de viajeros en España va a ser:

$$X_t = X_{t-12} + 225.635 + \varepsilon_t.$$

Para realizar las previsiones para un cierto mes se suma al mes del mismo nombre del año anterior el valor de c. Para calcular el error, en un cierto momento temporal, se haya la diferencia entre el valor de la serie y el estimado por el modelo.

Por ejemplo el valor previsto por el modelo para el mes numero 20 de la serie es:

$$X_{20} = X_8 + 225.635 = 6906.9 + 225.635 = 7132.\,5$$

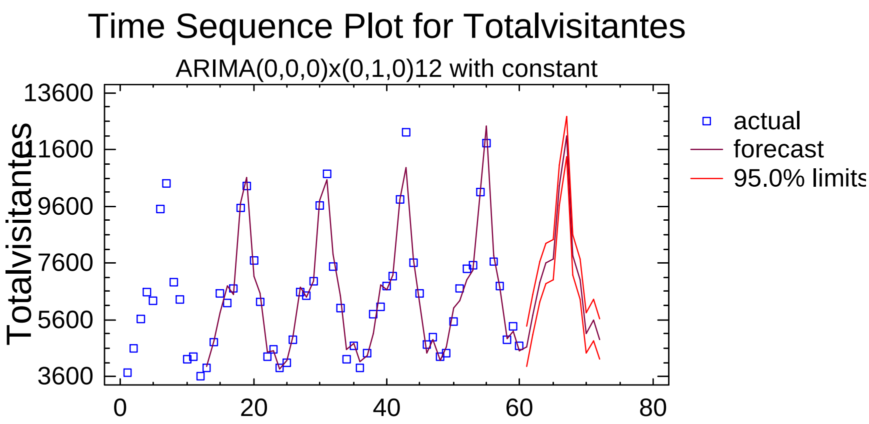
Como el valor de la serie en dicho periodo era 7688.8, resulta que el error de ajuste del modelo a la serie es

$$\varepsilon_{20} = 7688.8 - 7132.\,5 = 556.\,3$$

Ahora tendríamos que decidir si la serie residual es puramente aleatoria. En este caso se superan todos los tests de hipótesis realizados, así que se acepta que la componente residual es una serie de ruido blanco. En cuanto al valor de la desviación típica de esta variable de error es, para este modelo, algo mayor que para el modelo precedente, pues su valor es en este caso es 354.1 Si nos decidimos por este modelo para representar lan serie y lo empleamos para hacer prediciones para el mes de Marzo de 2004, esta previsión tomaría ahora el valor:

$$X_{62} = X_{50} + 225.635 = 5545.7 + 225.635 = 5771.\,3$$

En la figura siguiente están representados los valores de la serie primitiva, el valor calculado por el modelo para los cuatro últimos años dados en la tabla (el primer año se usa para inicializar la serie, por eso no hay valores predichos con el modelo para los puntos del primer año). Tambien están representados los valores previstos para el siguiente año con un intervalo de confianza al 95%.



Este modelo no resulta tan ajustado como el utilizado en 196, pues la varianza del error es mayor en este caso que en modelo clasico allí empleado. Se observa que en este caso los intervalos de confianza son más amplios, ya que el error tiene un mayor valor y además en este caso hay menos valores para estimar las predicciones, ya que no tenemos los datos correspondientes a los errores del primer año, pues perdemos estos 12 valores al pasar a la serie Y_t. Si queremos decidir cual de los dos modelos estudiados es más conveniente, la decisión no es del todo clara, pues tendríamos que elegir entre un modelo con menos error medio, como el primero, pero con algún problema de aceptación de uno de los tests de la variable residual, o por el segundo, que supera todos los tests a los que se ha sometido, que es más seguro desde el punto de vista teórico, pero que tiene la desventaja de que proporciona, por promedio, un mayor error. No hay ninguna respuesta clara a este dilema, pero cabe inclinarse por el segundo, pues el modelo cumple mejor las consideraciones teóricas, aunque tenga potencialmente más error.

Ejercicio 199 *Dada la serie temporal siguiente, cuyos datos están recogidos en la variable* plastic *del el fichero* ejemst.sf3*, estudia con un paquete estadístico la posibilidad de adaptación a los siguientes modelos: a) ARIMA(2,2,6) sin constante, b)ARIMA(3,2,4) sin constante, c) ARIMA(0,2,6) sin constante, d)ARIMA(0,1,3) con constante, e)ARIMA(0,1,1) sin constante.*
5000, 4965, 4496, 4491, 4566, 4585, 4724, 4951, 4917, 4888, 5087, 5082, 5039, 5054, 4940, 4913, 4871, 4901, 4864, 4750, 4856, 4959, 5004, 5415,

5550, 5657, 6010, 6109, 6052, 6391, 6798, 6740, 6778, 7005, 7045, 7279, 7367, 6934, 6506, 6374, 6066, 6102, 6204, 6138, 5938, 5781, 5813, 5811, 5818, 5982, 6132, 6111, 5948, 6056, 6342, 6626, 6591, 6302, 6132, 5837, 5572, 5744, 6005, 6239, 6523, 6652, 6585, 6622, 6754, 6712, 6675, 6882, 7011, 7140, 7197, 7411, 7233, 6958, 6960, 6927, 6814, 6757, 6765, 6870, 6954, 6551, 6022, 5974, 6052, 6033, 6030, 5944, 5543, 5416, 5571, 5571, 5627, 5679, 5455, 5443.

Seleccionar entre los cinco modelos el que creas reuna mejores características, explicando los motivos de esta elección detalla el modelo y halla la previsión para los primeros valores de la serie. Consideramos nulo el error del primer término.

Mostramos a continuación algunos de los resultados obtenidos con *Statgraphics plus 5.1* y los comentarios pertinentes*:*

Modelo	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	152.923	OK	OK	OK	OK	OK
(B)	170.76	OK	OK	OK	OK	OK
(C)	155.058	OK	OK	OK	OK	OK
(D)	158.095	OK	OK	OK	OK	OK
(E)	160.34	OK	OK	OK	OK	OK

Clave:

RMSE = Raiz Error Cuadrado Medio

RUNS = Test para excesivas ejecuciones arriba y abajo

RUNM = Test para excesivas ejecuciones por encima y por debajo de la mediana

AUTO = test de Box-Pierce para excesivas autocorrelaciones

MEAN = Test para la diferencia en la media de la 1^a mitad a la 2^a mitad

VAR = Test para la diferencia en la varianza en la 1^a mitad a la 2^a mitad

OK = no significativo (p >= 0.05)

Comentarios:

Todos los modelos tienen residuos que superan los test realizados para contrastar su concordancia con un proceso de ruido blanco, así que desde este punto de vista todos serían válidos. Rechazamos de momento el segundo, pues su error es algo mayor que los restantes, y decidiremos entre los otros en función de la significación de sus parámetros. La salida del programa al respecto se muestra a continuación. La significación de los parámetros puede medirse por medio del *p-value* asociado. Un *p-value* mayor que 0.05 indica que el parámetro al que se refiere no es significativo con una confianza del 95%. Entre dos modelos similares se optará por el más simple.

A continuación observamos los parámetros , la gráfica del ajuste y los correlogramas total y parcial de cada modelo

Modelo de pronóstico seleccionado: ARIMA(2,2,6)

Parámetro	Estimación	Error	Estd. t	P-Valor
AR(1)	0.00907164	0.00533039	1.70187	0.092232
AR(2)	-0.00966226	0.00466466	-2.07137	0.041187
MA(1)	0.294708	0.102059	2.88763	0.004861
MA(2)	0.73878	0.10067	7.33867	0.000000
MA(3)	-0.138771	0.12915	-1.0745	0.285472
MA(4)	0.134223	0.126142	1.06406	0.290149
MA(5)	0.173434	0.102796	1.68717	0.09503
MA(6)	-0.164909	0.10498	-1.57086	0.119724

En este modelo hay parámetros no significativos.Incluso no podemos afirmar el nivel 6 para la parte de medias móviles, ya que el coeficiente asociado a MA(6) no es significativo

Modelo de pronóstico seleccionado: ARIMA(0,2,6)

Parámetro	Estimación	Error	Estd. t	P-Valor
MA(1)	0.311148	0.101406	3.06835	0.002828
MA(2)	0.716899	0.100863	7.10766	0.000000
MA(3)	-0.161916	0.124747	-1.29795	0.197548
MA(4)	0.167026	0.123688	1.35038	0.180207
MA(5)	0.170065	0.101022	1.68344	0.095680
MA(6)	-0.206005	0.100292	-2.05407	0.042807

En este modelo tambien hay coeficientes que no son significativos

.

Modelo de pronóstico seleccionado: ARIMA(0,1,3) con constante

Parámetro	Estimación	Error	Estd. t	P-Valor
MA(1)	-0.676419	0.101952	-6.63469	0.000000
MA(2)	0.107473	0.119419	0.89997	0.370412
MA(3)	0.0484957	0.100194	0.4 84016	0.629488
Media	7.38729	24.5594	0.300792	0.764230
Constante	7.38729			

En este modelo el único coeficiente significativo es el que corresponde a MA(1).

Modelo de pronóstico seleccionado: ARIMA(0,1,1)

Parámetro	Estimación	Error	Estd. t	P-Valor
MA(1)	-0.676014 0	0802586	-8.42295	0.000000

El único parámetro de este modelo es significativo, el error previsto es casi igual que los restantes y es el más simple, así que nos inclinamos por este ultimo modelo para representar la serie dada.

Conclusion: El modelo seleccionado es ARIMA(0,1,1). Su expresión es $X_{t+1} = X_t + 0.676014\varepsilon_{t-1} + \varepsilon_t$.

Las previsiones para los primeros términos son:

t	X_t		$\widehat{X_t}$	$\widehat{\varepsilon_t}$
1	5000			
2	4965	5000+0.676014×0	5000.0	-35
3	4496	4965+0.676014× (-35)	4941. 3	445.3
4	4491	4496+0.676014× (-445.3)	4195. 0	296