

Estadística descriptiva: problemas propuestos

BENITO J. GONZÁLEZ RODRÍGUEZ (bjglez@ull.es)

DOMINGO HERNÁNDEZ ABREU (dhabreu@ull.es)

MATEO M. JIMÉNEZ PAIZ (mjimenez@ull.es)

M. ISABEL MARRERO RODRÍGUEZ (imarrero@ull.es)

ALEJANDRO SANABRIA GARCÍA (asgarcia@ull.es)

Departamento de Análisis Matemático
Universidad de La Laguna

Índice

6. Problemas propuestos	1
6.1. Variables no agrupadas	1
6.2. Variables agrupadas	2

ULL

Universidad
de La Laguna



6. Problemas propuestos

6.1. Variables no agrupadas

1. En una clínica infantil se ha ido anotando, durante un mes, el número de metros que cada niño anda, seguido y sin caerse, el primer día que comienza a caminar, obteniéndose la tabla de información adjunta:

número de metros	1	2	3	4	5	6	7	8
número de niños	2	6	10	5	10	3	2	2

Se pide:

- Tabla de frecuencias. Diagrama de barras para frecuencias absolutas, relativas y acumuladas.
- Mediana, media aritmética, moda y cuartiles.
- Varianza y desviación típica.
- ¿Entre qué valores se encuentra como mínimo el 75% de las observaciones?

Solución.

a)	x_i	n_i	N_i	f_i	F_i
	1	2	2	0.050	0.050
	2	6	8	0.150	0.200
	3	10	18	0.250	0.450
	4	5	23	0.125	0.575
	5	10	33	0.250	0.825
	6	3	36	0.075	0.900
	7	2	38	0.050	0.950
	8	2	40	0.050	1.000

b) $M_e = 4$, $\bar{x} = 4.1$, $M_{o_1} = 3$, $M_{o_2} = 5$, $P_{1/4} = 3$, $P_{3/4} = 5$; c) $\sigma^2 = 3.1$, $\sigma = 1.8$; d) (0.5, 7.6).

2. Las cifras dadas en la tabla adjunta corresponden a miligramos de hidroxiprolina absorbidos por 1 miligramo de masa intestinal, analizados en distintos pacientes:

miligramos de hidroxiprolina	77.3	61.2	82.4	75.9	61	70.2	65	80
número de pacientes	3	10	15	13	8	5	2	0

Se pide:

- Confeccionar la tabla de frecuencias.

- b) Calcular la media, mediana, moda y cuartiles.
- c) Calcular la desviación típica de la muestra.
- d) ¿Qué porcentaje de observaciones se concentra en el intervalo $(\bar{x} - 5\sigma, \bar{x} + 5\sigma)$?

Solución.

a)	x_i	n_i	N_i	f_i (%)	F_i (%)
	61	8	8	14.3	14.3
	61.2	10	18	17.9	32.2
	65	2	20	3.6	35.8
	70.2	5	25	8.9	44.7
	75.9	13	38	23.2	67.9
	77.3	3	41	5.4	73.3
	82.4	15	56	26.8	100.0

b) $\bar{x} = 72.1$, $M_e = 75.9$, $M_o = 82.4$, $P_{1/4} = 61.2$,
 $P_{3/4} = 82.4$; c) $\sigma = 8.6$; d) un mínimo del 96 %.

6.2. Variables agrupadas

3. Los valores del pH sanguíneo de 32 individuos son los siguientes:

7.33 7.31 7.26 7.33 7.37 7.27 7.30 7.33
 7.33 7.32 7.35 7.39 7.33 7.38 7.33 7.31
 7.37 7.35 7.34 7.32 7.29 7.35 7.38 7.32
 7.32 7.33 7.32 7.40 7.33 7.32 7.34 7.33

- a) Agrupar los datos en 5 intervalos y confeccionar la tabla de frecuencias.
- b) Determinar la media aritmética, la moda y la mediana.
- c) Determinar el tercer decil.
- d) Determinar el porcentaje de individuos que se encuentra fuera del intervalo $(\bar{x} - 4\sigma, \bar{x} + 4\sigma)$.

Solución.

a)	IC	x_i	n_i	N_i	f_i (%)	F_i (%)	
	[7.26,7.29)	7.275	2	2	6.250	6.250	b) $\bar{x} = 7.34$, intervalo modal: [7.32,7.35), $M_e = 7.34$; c) $D_3 = 7.33$; d) a lo sumo el 6.25%.
	[7.29,7.32)	7.305	4	6	12.500	18.750	
	[7.32,7.35)	7.335	17	23	53.125	71.875	
	[7.35,7.38)	7.365	5	28	15.625	87.500	
	[7.38,7.41)	7.395	4	32	12.500	100.000	

4. En pacientes con distrofia muscular progresiva (*enfermedad de Duchenne*), la actividad de creatinquinasa sérica se eleva llamativamente sobre el valor normal de 50 unidades por litro. Los siguientes datos son niveles séricos de creatinquinasa medidos en 24 jóvenes pacientes con la enfermedad confirmada (en unidades por litro):

3720	3795	3340	5600	3800	3580
5500	2000	1570	2360	1500	1840
3725	3790	3345	3805	5595	3575
1995	5505	2055	1575	1835	1505

Se pide:

- a) Agrupar los datos en 5 intervalos de clase.
- b) Determinar la media y la desviación típica. Calcular la moda y la mediana.
- c) Hallar el tercer cuartil, el séptimo decil y el centil 25.

Solución.

a)	IC	x_i	n_i	N_i	
	[1500,2321)	1910.5	9	9	b) $\bar{x} = 3176.2$, $\sigma = 1160.6$, intervalo modal: [3142,3963), $M_e = 3306.2$; c) $P_{3/4} = 3798.8$, $D_7 = 3700.3$, $P_{25} = 2047.3$.
	[2321,3142)	2731.5	1	10	
	[3142,3963)	3552.5	10	20	
	[3963,4784)	4373.5	0	20	
	[4784,5605)	5194.5	4	24	

5. Se ha medido el metabolismo basal de 25 individuos, resultando los siguientes datos:

102 98 93 100 98
 115 130 100 86 95
 116 118 88 102 128
 112 114 106 114 100
 120 106 110 100 106

Se pide:

- Agrupar los datos en 4 intervalos de clase.
- Determinar la media y la desviación típica. Calcular la moda y la mediana.
- Hallar el segundo cuartil y el centil 40.

Solución.

a)	IC	x_i	n_i	N_i
	[86, 98)	92	4	4
	[98, 110)	104	11	15
	[110, 122)	116	8	23
	[122, 134)	128	2	25

b) $\bar{x} = 107.8$, $\sigma = 10.0$, intervalo modal:
 [98, 110), $M_e = 107.3$; c) $P_{2/4} = M_e = 107.3$,
 $P_{40} = 104.6$.

6. En el estudio de pautas de crecimiento en niños, una variable importante es la edad del niño cuando comienza el crecimiento rápido de la adolescencia. Las siguientes observaciones fueron obtenidas en un estudio de 35 chicos y 40 chicas (edad en años):

chicos						
16.0	14.9	14.1	14.8	14.4	14.0	14.6
15.2	14.7	13.6	14.6	16.1	13.2	13.2
14.9	14.1	15.4	15.3	14.4	14.8	14.8
13.5	15.1	13.5	15.0	14.6	15.4	15.9
13.7	15.9	14.7	14.5	14.4	13.8	15.3

chicas						
12.2	13.7	13.3	12.3	12.5	12.9	11.9
13.4	12.4	12.6	13.5	12.5	13.4	11.7
13.7	12.1	14.1	11.8	12.8	12.9	11.6
13.1	13.3	13.5	14.7	12.3	11.6	13.1
12.7	12.7	12.0	11.4	13.5	12.4	12.1
11.6	13.5	14.3	12.6	12.1		

Para cada grupo, se pide:

- a) Agrupar el conjunto de datos en seis intervalos de clase.
- b) Representar mediante histogramas las frecuencias absolutas, las relativas y las acumuladas.
- c) Determinar la media aritmética, la moda y la mediana.
- d) Hallar los cuartiles.
- e) Calcular la desviación típica.

Solución.

chicos						
a)	IC	x_i	n_i	N_i	f_i	F_i
	[13.2, 13.7)	13.45	5	5	0.14	0.14
	[13.7, 14.2)	13.95	5	10	0.14	0.28
	[14.2, 14.7)	14.45	7	17	0.20	0.48
	[14.7, 15.2)	14.95	9	26	0.26	0.74
	[15.2, 15.7)	15.45	5	31	0.14	0.88
	[15.7, 16.2)	15.95	4	35	0.11	0.99

c) $\bar{x} = 14.7$, intervalo modal: [14.7, 15.2), $M_e = 14.7$; d) $P_{1/4} = 14.1$, $P_{3/4} = 15.2$; e) $\sigma = 0.7$.

chicas							
a)	IC	x_i	n_i	N_i	f_i	F_i	
	[11.4,12.0)	11.7	7	7	0.175	0.175	c) $\bar{x} = 12.8$, intervalo modal: [12.0,12.6), $M_e = 12.7$; d) $P_{1/4} = 12.2$, $P_{3/4} = 13.4$; e) $\sigma = 0.8$.
	[12.0,12.6)	12.3	11	18	0.275	0.450	
	[12.6,13.2)	12.9	9	27	0.225	0.675	
	[13.2,13.8)	13.5	10	37	0.250	0.925	
	[13.8,14.4)	14.1	2	39	0.050	0.975	
	[14.4,15.0)	14.7	1	40	0.025	1.000	

7. Para investigar la distribución de los diámetros de los ejes de acero producidos en un proceso de laminación se tomaron las siguientes 20 mediciones:

2.510 2.517 2.522 2.522 2.510 2.511 2.511
 2.519 2.543 2.525 2.532 2.527 2.536 2.505
 2.541 2.512 2.515 2.521 2.536 2.529

Se pide:

- a) Agrupar los datos en 5 intervalos de clase.
- b) Calcular la media y la desviación típica.
- c) Calcular la mediana, el tercer decil y el centil 65.

Solución.

a)	IC	x_i	n_i	N_i	
	[2.505,2.513)	2.509	6	6	b) $\bar{x} = 2.522$, $\sigma = 0.011$; c) $M_e = 2.523$, $D_3 = 2.513$, $P_{65} = 2.527$.
	[2.513,2.521)	2.517	3	9	
	[2.521,2.529)	2.525	5	14	
	[2.529,2.537)	2.533	4	18	
	[2.537,2.545)	2.541	2	20	

8. Se realiza un estudio sobre la edad de las mujeres que utilizan anticonceptivos orales, obteniéndose los siguientes datos:

intervalo de clase	frecuencia absoluta
[14.5,19.5)	171
[19.5,24.5)	785
[24.5,29.5)	837
[29.5,34.5)	554
[34.5,39.5)	382
[39.5,44.5)	432
[44.5,49.5)	562
[49.5,54.5)	610
[54.5,59.5)	490
[59.5,64.5)	258
[64.5,69.5)	153
[69.5,74.5)	60

- Obtener las frecuencias relativas y las correspondientes acumuladas.
- Obtener las marcas de clase.
- Calcular la media, la mediana y la moda.
- Calcular la desviación típica.
- Calcular los cuartiles, el decil 4 y el centil 32.

Solución.

a), b)	IC	x_i	n_i	N_i	f_i	F_i
	[14.5, 19.5)	17	171	171	0.032	0.032
	[19.5, 24.5)	22	785	956	0.148	0.180
	[24.5, 29.5)	27	837	1793	0.158	0.338
	[29.5, 34.5)	32	554	2347	0.105	0.443
	[34.5, 39.5)	37	382	2729	0.072	0.515
	[39.5, 44.5)	42	432	3161	0.082	0.597
	[44.5, 49.5)	47	562	3723	0.106	0.703
	[49.5, 54.5)	52	610	4333	0.115	0.818
	[54.5, 59.5)	57	490	4823	0.093	0.911
	[59.5, 64.5)	62	258	5081	0.049	0.960
	[64.5, 69.5)	67	153	5234	0.029	0.989
	[69.5, 74.5)	72	60	5294	0.011	1.000

c) $\bar{x} = 39.6$, $M_e = 38.4$, intervalo modal: [24.5, 29.5); d) $\sigma = 14.3$; e) $P_{1/4} = 26.7$, $P_{3/4} = 51.5$, $D_4 = 32.4$, $P_{32} = 28.9$.

9. Se realizó un estudio de la *enfermedad de Hodgkin* (enfermedad sistémica, de curso progresivo, que afecta predominantemente a los ganglios linfáticos y al bazo), restringido a pacientes con edades inferiores a 40 años. Un propósito del estudio fue comparar la distribución de casos por edad de hombres con los de mujeres. Los resultados son los siguientes:

hombres		mujeres	
intervalos de clase (años)	frecuencia	intervalos de clase (años)	frecuencia
[4.5,9.5)	1	[4.5,9.5)	0
[9.5,14.5)	4	[9.5,14.5)	2
[14.5,19.5)	7	[14.5,19.5)	10
[19.5,24.5)	23	[19.5,24.5)	7
[24.5,29.5)	16	[24.5,29.5)	3
[29.5,34.5)	7	[29.5,34.5)	5
[34.5,39.5)	10	[34.5,39.5)	2

- a) Obtener las frecuencias relativas y las correspondientes acumuladas. Representar mediante histogramas las frecuencias absolutas y las acumuladas absolutas de los dos conjuntos de datos.
- b) Obtener las marcas de clase.
- c) Calcular la media, la mediana y la moda.
- d) Calcular la desviación típica.
- e) Calcular los cuartiles, el decil 8 y el centil 49.

Solución.

hombres						
a), b)	IC	x_i	n_i	N_i	f_i	F_i
	[4.5,9.5)	7	1	1	0.015	0.015
	[9.5,14.5)	12	4	5	0.059	0.074
	[14.5,19.5)	17	7	12	0.103	0.177
	[19.5,24.5)	22	23	35	0.338	0.515
	[24.5,29.5)	27	16	51	0.235	0.750
	[29.5,34.5)	32	7	58	0.103	0.853
	[34.5,39.5)	37	10	68	0.147	1.000

c) $\bar{x} = 25.1$, $M_e = 24.3$, intervalo modal: [19.5,24.5); d) $\sigma = 7.2$; e) $P_{1/4} = 20.6$, $P_{3/4} = 29.5$, $D_8 = 41.6$, $P_{49} = 24.1$.

		mujeres					
<i>a), b)</i>	IC	x_i	n_i	N_i	f_i	F_i	
	[4.5, 9.5)	7	0	0	0.000	0.000	<i>c)</i> $\bar{x} = 22.9$, $M_e = 21.3$, intervalo modal: [14.5, 19.5); <i>d)</i> $\sigma = 7.1$; <i>e)</i> $P_{1/4} = 17.1$, $P_{3/4} = 29.1$, $D_8 = 30.7$, $P_{49} = 21.1$.
	[9.5, 14.5)	12	2	2	0.069	0.069	
	[14.5, 19.5)	17	10	12	0.345	0.414	
	[19.5, 24.5)	22	7	19	0.241	0.655	
	[24.5, 29.5)	27	3	22	0.103	0.758	
	[29.5, 34.5)	32	5	27	0.172	0.930	
	[34.5, 39.5)	37	2	29	0.069	1.000	

10. Para establecer la importancia del pinzamiento precoz o tardío del cordón umbilical sobre el volumen sanguíneo del recién nacido se han estudiado los volúmenes sanguíneos (expresados en porcentajes de peso corporal) de 20 niños a los que se les pinzó el cordón precozmente. Los resultados obtenidos son:

12.7 9.2 8.4 8.7 9.4 10.3 11.2 10.6 8.3 9.6
9.2 9.1 10.3 11.2 9.7 9.4 10.1 11.8 9.7 9.8

Se pide:

- a)* Agrupar los datos en 5 intervalos.
b) Calcular la media, desviación típica, moda, mediana, primer cuartil, séptimo decil y centil 33.

Solución.

<i>a)</i>	IC	x_i	n_i	N_i	f_i	F_i	
	[8.3, 9.2)	8.75	4	4	0.20	0.20	<i>b)</i> $\bar{x} = 10.1$, $\sigma = 1.0$, intervalo modal: [9.2, 10.1), $M_e = 9.9$, $P_{1/4} = 9.3$, $D_7 = 0.3$, $P_{33} = 9.5$.
	[9.2, 10.1)	9.65	8	12	0.40	0.60	
	[10.1, 11.0)	10.55	4	16	0.20	0.80	
	[11.0, 11.9)	11.45	3	19	0.15	0.95	
	[11.9, 12.8)	12.35	1	20	0.05	1.00	

11. Sea X una variable estadística. Se conoce que su media es 100 y su varianza 25. ¿Qué porcentaje de observaciones se encuentra fuera del intervalo (85, 115)?

Solución. A lo sumo el 11.1%.