

## Problemas de Estadística Descriptiva. Resumen.

1. Un aeropuerto importante contrató recientemente al consultor Manuel López para estudiar el problema de los retrasos en el tráfico aéreo. En la siguiente tabla se puede observar la cantidad de minutos que los aviones llegaron tarde en una muestra de vuelos:

Minutos de retraso	0<10	10<20	20<30	30<40	40<50	50<60
Número de vuelos	30	25	13	6	5	4

- a. Estime el número medio de minutos de retraso.
- b. Estime la varianza y la desviación estándar de la muestra.

### Solución

- a. Para calcular el número medio de minutos de retraso, construimos la tabla siguiente donde la última fila en rojo hay las sumas las frecuencias y del producto de las frecuencias por los valores centrales:

Minutos de retraso	Frecuencia $f_i$ (número de vuelos)	Valores centrales $m_i$	$f_i \cdot m_i$
0<10	30	5	150
10<20	25	15	375
20<30	13	25	325
30<40	6	35	210
40<50	5	45	225
50<60	4	55	220
<b>Sumas</b>	<b>83</b>		<b>1505</b>

El número medio de minutos de retraso será:

$$\bar{x} = \frac{1505}{83} = 18.133.$$

- b. Para hallar la varianza y desviación típica, añadimos las columnas siguientes a la tabla anterior:

Minutos	$f_i$	$m_i$	$f_i \cdot m_i$	$m_i - \bar{x}$	$(m_i - \bar{x})^2$	$f_i \cdot (m_i - \bar{x})^2$
0<10	30	5	150	-13.133	172.475689	5173.9
10<20	25	15	375	-3.133	9.815689	245.319
20<30	13	25	325	6.867	47.155689	613.108
30<40	6	35	210	16.867	284.495689	1707.069
40<50	5	45	225	26.867	721.835689	3609.305
50<60	4	55	220	36.867	1359.175689	5436.841
<b>Sumas</b>	<b>83</b>		<b>1505</b>			<b>16785.542</b>

La varianza será:

$$s_x^2 = \frac{16785.542}{83 - 1} = \frac{16785.542}{82} = 204.702.$$

La desviación típica será:

$$s_x = \sqrt{s_x^2} = \sqrt{204.701734} = 14.307.$$

2. Considere las siguientes cuatro poblaciones:

- 1, 2, 3, 4, 5, 6, 7, 8
- 1, 1, 1, 1, 8, 8, 8, 8
- 1, 1, 4, 4, 5, 5, 8, 8
- -6, -3, 0, 3, 6, 9, 12, 15

Todas estas poblaciones tienen la misma media. Sin hacer los cálculos, organice las poblaciones de acuerdo con las magnitudes de sus varianzas, de menor a mayor. Luego calcule cada una de las variaciones manualmente.

## Solución

La media común de las 4 poblaciones vale 4.5.

Observamos que están ordenadas. La población con menos varianza estará entre la primera y la tercera que es dónde las diferencias con la media es más pequeña. Sin embargo, en la tercera es donde habrá diferencias más acentuadas ya que los valores extremos 1 y 8 se repiten más. En resumen las poblaciones con menos varianza serían: la primera y seguidamente la tercera.

Luego quedan la segunda y la cuarta. Es claro que el rango de valores en la cuarta es mucho mayor que en la segunda. Por tanto, las poblaciones que vienen en orden de menos a más varianza serían la segunda y por último la cuarta.

Calculemos seguidamente las varianzas a mano:

- Primera población:

$x_i$	$x_i - 4.5$	$(x_i - 4.5)^2$
1	-3.5	12.25
2	-2.5	6.25
3	-1.5	2.25
4	-0.5	0.25
5	0.5	0.25
6	1.5	2.25
7	2.5	6.25
8	3.5	12.25
<b>Sumas</b>		<b>42</b>

La varianza de la primera población será:  $s_{x,1}^2 = \frac{42}{7} = 6$ .

- Segunda población:

$x_i$	$x_i - 4.5$	$(x_i - 4.5)^2$
1	-3.5	12.25
1	-3.5	12.25
1	-3.5	12.25
1	-3.5	12.25
8	3.5	12.25
8	3.5	12.25
8	3.5	12.25
8	3.5	12.25
<b>Sumas</b>		<b>98</b>

La varianza de la primera población será:  $s_{x,2}^2 = \frac{98}{7} = 14$ .

- Tercera población:

$x_i$	$x_i - 4.5$	$(x_i - 4.5)^2$
1	-3.5	12.25
1	-3.5	12.25
4	-0.5	0.25
4	-0.5	0.25
5	0.5	0.25
5	0.5	0.25
8	3.5	12.25
8	3.5	12.25
<b>Sumas</b>		<b>50</b>

La varianza de la primera población será:  $s_{x,3}^2 = \frac{50}{7} = 7.1428571$ .

- Cuarta población:

$x_i$	$x_i - 4.5$	$(x_i - 4.5)^2$
-6	-10.5	110.25
-3	-7.5	56.25
0	-4.5	20.25
3	-1.5	2.25
6	1.5	2.25
9	4.5	20.25
12	7.5	56.25
15	10.5	110.25
<b>Sumas</b>		<b>378</b>

La varianza de la primera población será:  $s_{x,4}^2 = \frac{378}{7} = 54$ .

Por tanto, se cumple, tal como vaticinamos:

$$s_{x,1}^2 < s_{x,3}^2 < s_{x,2}^2 < s_{x,4}^2.$$

3. Los tiempos en minutos que están 50 clientes en un supermercado local para realizar la compra son los siguientes:

26.88	28.60	20.73	34.00	35.87	25.99	20.94	26.45	29.54	26.27
26.51	29.70	29.55	33.52	30.49	31.49	21.28	23.57	22.47	23.15
19.51	23.85	30.98	9.81	26.59	29.68	30.48	25.38	23.49	25.11
19.35	33.80	23.14	13.56	24.63	24.26	37.18	22.20	21.37	28.30
11.02	25.59	24.38	25.29	29.17	25.55	26.94	27.24	19.10	27.44

- Calcular el tiempo medio que tardan los 50 clientes.
- Calcular la varianza y la desviación típica de los tiempos.
- Calcular el percentil 90.
- Calcular los 5 números resumen.
- Calcular el coeficiente de variación.

### Solución:

- a. El tiempo medio será:

$$\bar{x} = \frac{26.88 + \dots + 27.44}{50} = \frac{1281.39}{50} = 25.6278.$$

- b. La varianza de los tiempos será:

$$s_x^2 = \left( \frac{26.88^2 + \dots + 27.44^2}{50} - 25.6278^2 \right) \cdot \frac{50}{49} = \left( \frac{34353.7941}{50} - 25.6278^2 \right) \cdot \frac{50}{49} = 30.9099481.$$

- c. Para calcular el percentil 90, primero ordenamos los tiempos:

9.81	11.02	13.56	19.10	19.35	19.51	20.73	20.94	21.28	21.37
22.20	22.47	23.14	23.15	23.49	23.57	23.85	24.26	24.38	24.63
25.11	25.29	25.38	25.55	25.59	25.99	26.27	26.45	26.51	26.59
26.88	26.94	27.24	27.44	28.30	28.60	29.17	29.54	29.55	29.68
29.70	30.48	30.49	30.98	31.49	33.52	33.80	34.00	35.87	37.18

El percentil 90 es el que deja a su “izquierda” el 90% de los valores. Por tanto, deja a su izquierda  $0.9 \cdot 50 = 45$  valores. El percentil 90 será el valor que ocupa el lugar 45 en los tiempos ordenados. Dicho valor vale 31.49.

- d. Los cinco números resumen son el mínimo, el primer cuartil, la mediana, el segundo cuartil y el máximo:

- el mínimo vale: 9.81.
- el primer cuartil es el que deja a su izquierda el 25% de los datos, es decir, es la mediana de los 25 primeros datos que correspondería al valor que ocupa el lugar 13. Dicho valor vale:  $Q_1 = 23.14$ .
- la mediana es el que deja a su izquierda el 50% de los datos. Correspondería a la semisuma de los dos valores centrales que ocupan los lugares 25 y 26. Dichos valores son 25.59 y 25.99. La mediana será, pues,  $Q_2 = \frac{25.59+25.99}{2} = 25.79$ .
- el tercer cuartil es el que deja a su izquierda el 75% de los datos, es decir, es la mediana de los 25 últimos datos que correspondería al valor que ocupa el lugar  $25 + 13 = 38$ . Dicho valor vale:  $Q_3 = 29.54$ .
- el máximo vale: 37.18.

4. La tabla siguiente nos da unos indicadores socio-económicos para cada una de las 47 provincias de habla francesa de Suiza en 1888:

	Fertilidad	Agricultura	Examen	Educación	Católicos	Mortalidad infantil
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4
Veveyse	87.1	64.5	14	6	98.61	24.5
Aigle	64.1	62.0	21	12	8.52	16.5
Aubonne	66.9	67.5	14	7	2.27	19.1
Avenches	68.9	60.7	19	12	4.43	22.7
Cossonay	61.7	69.3	22	5	2.82	18.7
Echallens	68.3	72.6	18	2	24.20	21.2
Grandson	71.7	34.0	17	8	3.30	20.0
Lausanne	55.7	19.4	26	28	12.11	20.2
La Vallee	54.3	15.2	31	20	2.15	10.8
Lavaux	65.1	73.0	19	9	2.84	20.0
Morges	65.5	59.8	22	10	5.23	18.0
Moudon	65.0	55.1	14	3	4.52	22.4
Nyone	56.6	50.9	22	12	15.14	16.7
Orbe	57.4	54.1	20	6	4.20	15.3
Oron	72.5	71.2	12	1	2.40	21.0
Payerne	74.2	58.1	14	8	5.23	23.8
Paysd'enhaut	72.0	63.5	6	3	2.56	18.0
Rolle	60.5	60.8	16	10	7.72	16.3
Vevey	58.3	26.8	25	19	18.46	20.9
Yverdon	65.4	49.5	15	8	6.10	22.5
Conthey	75.5	85.9	3	2	99.71	15.1
Entremont	69.3	84.9	7	6	99.68	19.8
Herens	77.3	89.7	5	2	100.00	18.3
Martigwy	70.5	78.2	12	6	98.96	19.4
Monthey	79.4	64.9	7	3	98.22	20.2
St Maurice	65.0	75.9	9	9	99.06	17.8
Sierre	92.2	84.6	3	3	99.46	16.3
Sion	79.3	63.1	13	13	96.83	18.1
Boudry	70.4	38.4	26	12	5.62	20.3
La Chauxdfnd	65.7	7.7	29	11	13.79	20.5
Le Locle	72.7	16.7	22	13	11.22	18.9
Neuchatel	64.4	17.6	35	32	16.92	23.0
Val de Ruz	77.6	37.6	15	7	4.97	20.0
ValdeTravers	67.6	18.7	25	7	8.65	19.5
V. De Geneve	35.0	1.2	37	53	42.34	18.0
Rive Droite	44.7	46.6	16	29	50.43	18.2
Rive Gauche	42.8	27.7	22	29	58.33	19.3

donde:

- Fertilidad: indica el índice de fertilidad de la provincia,
  - Agricultura: indica el porcentaje de hombres que se dedican a la agricultura,
  - Examen: indica el porcentaje de reclutas que reciben la calificación más alta en el examen del ejército,
  - Educación: indica el porcentaje de reclutas que tienen una educación superior a la primaria,
  - Católicos: indica el porcentaje de católicos,
  - Mortalidad infantil: indica el porcentaje de bebés que viven menos de un año.
- a. Dar la tabla de frecuencias de la variable Educación.
- b. Calcula la media y la varianza de la variable Educación como datos agrupados.
- c. Halla la media ponderada de la variable Fertilidad considerando como pesos los valores de la variable educación.

## Solución

- a. La tabla de frecuencias de la variable Educación es la siguiente:

Educación	Frecuencia $f_i$
1	1
2	3
3	4
5	2
6	4
7	7
8	4
9	3
10	2
11	1
12	5
13	3
15	1
19	1
20	1
28	1
29	2
32	1
53	1

- b. Para calcular la media de la variable Educación, construimos la tabla siguiente:

Valores educación $x_i$	$f_i$	$f_i \cdot x_i$
1	1	1
2	3	6
3	4	12
5	2	10
6	4	24
7	7	49
8	4	32
9	3	27
10	2	20
11	1	11
12	5	60
13	3	39
15	1	15
19	1	19
20	1	20
28	1	28
29	2	58
32	1	32
53	1	53
<b>Sumas</b>	<b>47</b>	<b>516</b>

La media será:

$$\bar{x} = \frac{516}{47} = 10.978723.$$

c. Para calcular la varianza, hacemos la tabla siguiente:

Valores educación $x_i$	$f_i$	$f_i \cdot x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$
1	1	1	-9.979	99.575	99.575
2	3	6	-8.979	80.617	241.852
3	4	12	-7.979	63.66	254.64
5	2	10	-5.979	35.745	71.49
6	4	24	-4.979	24.788	99.151
7	7	49	-3.979	15.83	110.812
8	4	32	-2.979	8.873	35.491
9	3	27	-1.979	3.915	11.746
10	2	20	-0.979	0.958	1.916
11	1	11	0.021	0	0
12	5	60	1.021	1.043	5.215
13	3	39	2.021	4.086	12.257
15	1	15	4.021	16.171	16.171
19	1	19	8.021	64.341	64.341
20	1	20	9.021	81.383	81.383
28	1	28	17.021	289.724	289.724
29	2	58	18.021	324.766	649.533
32	1	32	21.021	441.894	441.894
53	1	53	42.021	1765.788	1765.788
<b>Sumas</b>	<b>47</b>	<b>516</b>			<b>4252.979</b>

La varianza y la desviación típica serán, pues,

$$s_x^2 = \frac{4252.979}{47 - 1} = 92.456059, \quad s_x = \sqrt{92.456059} = 9.615407.$$



5. Usando los datos de la tabla anterior, queremos estudiar la posible relación entre las variables Fertilidad y Examen.
- Realiza un gráfico de puntos de las variables anteriores, indicando en el eje  $X$  o abscisas la variable Examen y en el  $Y$ , la variable *Fertilidad*.
  - Calcula la covarianza y la correlación entre las variables anteriores. ¿A qué conclusión llegas?