

ESTADÍSTICA DESCRIPTIVA Y ANÁLISIS EXPLORATORIO DE DATOS

Resuelva los siguientes problemas, justificando convenientemente.

- 1) A continuación aparecen algunas conclusiones extraídas de una publicación. En cada una de ellas identifique la variable X que se mide y la medida descriptiva a la que hace referencia.
 - a) \$ 10,41 es el mayor pago por hora a vendedores en la industria de materiales de construcción de la Argentina durante el año 2012 que obtiene el 50% que menos cobra.
 - b) 69% de los trabajadores estadounidenses de 16 años o mayores, durante el año 2011, trabajan al menos 40 horas por semana.
 - c) 70% de todos los profesores auxiliares en EEUU, durante los últimos 20 años, ganan anualmente menos de 91823 dólares.

- 2) La empresa "Golosos" tiene necesidad de conocer algunos detalles de su producción y venta. Algunas preguntas que se hace son las que siguen:
 - a) ¿Es cierto que la demanda mensual de chicles está creciendo?
 - b) En el mes anterior ¿qué golosina se vendió más?
 - c) ¿Es cierto que se espera que a mayor demanda diaria de chicles haya menor demanda de pastillas?Ud. es empleado de esta firma y debe responder a esas preguntas mediante gráficos, ¿cuáles utilizaría y por qué? Explique cómo detectaría en ellos la respuesta correspondiente.

- 3) Proponga y justifique la elección de:
 - a) un lote de datos en el que más de la mitad sean iguales a la mediana del lote.
 - b) una distribución de frecuencias que sea simétrica y en la que la media aritmética y el modo no sean iguales.
 - c) un lote de 16 datos numéricos que sean distintos y cuya media aritmética sea exactamente igual a cero.

- 4) Dibuje los histogramas correspondientes a dos distribuciones de frecuencias con los mismos intervalos de clases, en donde ambos lotes sean de igual magnitud y además tengan igual promedio, pero diferente variabilidad alrededor del mismo.

- 5) La Universidad Nacional de Cuyo está analizando los resultados obtenidos por los ingresantes 2014 en los exámenes de las distintas áreas en donde han sido evaluados. En el área matemática se observó una calificación promedio de 480 puntos con una variabilidad promedio de 100 puntos, si además el coordinador del área matemática detectó que la distribución de estas calificaciones es simétrica y tiene forma de campana:
 - a) ¿Cuál es el porcentaje aproximado de calificaciones de matemática que estuvieron entre 380 y 580 puntos?
 - b) ¿Cuál es el porcentaje aproximado de calificaciones de matemática que estuvieron por arriba de 680 puntos?
 - c) Carlos obtuvo 580 puntos en el examen de matemática, ¿Cuál es el porcentaje aproximado de estudiantes que obtuvieron una calificación, en matemática, inferior a la de Carlos?
 - d) Lucía obtuvo 680 puntos en el examen de matemática, ¿Cuál es el porcentaje aproximado de estudiantes que tuvieron una calificación, en matemática, más alta que Carlos pero más baja que Lucía?

6) Después de examinar los registros de facturación mensuales de la compañía R&S, ubicada en la ciudad de Santa Rosa (La Pampa), el auditor toma una muestra de 20 saldos adeudados (en miles de \$) al 20 de diciembre de 2013. Los mismos se muestran a continuación:

18 11 7 5 10 33 9 12 4 7
11 10 6 37 15 18 10 21 3 26

- a) Indique la variable bajo estudio.
- b) Las cantidades adeudadas a la compañía, ¿muestran una marcada dispersión alrededor de la media? Justifique.
- c) Realice un gráfico (el más conveniente) que muestre cómo se distribuye la variable bajo estudio. Justifique la elección del mismo.
- d) ¿A qué conclusiones llega?

7) Dada la siguiente distribución de frecuencias:

**Distribución de edades de los operarios de la empresa Águila S.A.
Rosario (Sta. Fe) - octubre 2013 -**

Edad de los operarios	Nº de operarios
[20, 30)	4
[30, 40)	32
[40, 50)	28
[50, 60)	6

Fuente: Dpto. Personal, empresa Águila S.A.

Responda:

- a) ¿Se aplicó la regla empírica de Sturges para su construcción?
- b) ¿Es una distribución simétrica?
- c) ¿La mayoría de los operarios tienen de 30 a menos de 40 años?
- d) ¿Es posible representar esta distribución a través de un box-plot? De ser posible realícelo.
- e) ¿Qué medida elegiría para representar la tendencia central de este lote? ¿Cuál para su variabilidad?

8) Una compañía aplica dos métodos diferentes para surtir los pedidos de sus clientes. Al estudiar los mismos encontró que las distribuciones del tiempo de entrega según el método utilizado tienen algunas diferencias entre sí. Si bien ambas, A y B, tienen forma acampanada, se da que la menor dispersión respecto a su promedio, el menor tiempo promedio y el menor rango se presentan si el método es el A. Realice un esquema de la situación y en base a ello explique qué método recomendaría usted y porqué.

9) Los siguientes datos son los valores unitarios, en dólares, de las acciones de 30 compañías cotizando en la Bolsa de Buenos Aires obtenidos en el mes de noviembre de 2013 por la consultora P&S:

70 92 81 62 69 63 99 69 97 197
82 53 65 86 73 65 77 89 110 63
102 68 112 75 68 76 75 100 120 72

- a) Defina y clasifique la variable que se estudia.
- b) Construya:
 - * el diagrama de tallo y hoja
 - * un box-plot
- c) Calcule (sin agrupar): media, mediana, cuartiles, rango, coeficiente de variación.

- d) Analice la simetría del conjunto de datos.
- e) ¿Qué porcentaje de los datos está contenido entre ± 1 desviaciones estándares de la media?
- f) Ud. trabaja en la consultora P&S y es el responsable del informe que se entregará a la firma que solicitó dicho análisis. Redacte el mismo.
- g) Repita en inciso b, cambiando al dato 197 por 187.

10) Los siguientes datos ordenados corresponden a los sueldos mensuales (en \$) de un grupo de empleados -Junio de 2013- seleccionados al azar por el Departamento Contable de la industria PSP ubicada en ciudad de Salta, República Argentina:

7590	7695	7742	7869	7999	8780	8790	8890
9830	9900	10190	10250	10523	10639	10760	11598
11650	11785	11800	11870	11940	12012	12015	12300

Analice el comportamiento de estos sueldos realizando un box-plot y un diagrama de tallo y hoja.

11) En una empresa se ha clasificado al total de empleados según su salario, obteniéndose la siguiente distribución:

**Distribución de los salarios mensuales de empleados de CARRO'S
Paraná (Entre Ríos) – febrero 2013-**

Salario (en pesos)	Cantidad de Empleados
[500 , 1000)	4
[1000 , 1500)	9
[1500 , 2000)	16
[2000 , 2500)	24
[2500 , 3000)	27
[3000 , 3500)	14
[3500 , 4000)	6

Fuente: Dpto. Administración Contable – CARRO'S-

Si se decide aumentar los plus por horas extras a los empleados cuyo salario se encuentre por encima del 25 % del menor salario y por debajo del 32 % del mayor salario, aproximadamente:

- a) ¿entre qué valores se encuentran los salarios de los empleados que recibirán un aumento en el plus por hora extra?
- b) ¿qué cantidad de empleados contarán con dicho beneficio?
- c) Si se le incrementan los días de vacaciones a aquellos empleados que cobran más de $\mu + \sigma$, aproximadamente ¿qué cantidad de empleados tendrán más días de vacaciones?

12) Para determinar el número de cajas de pago que en el futuro es necesario tener, la cadena de supermercados *SIEMPRE*, que se desarrolla en la zona pampeana de la República Argentina, cree que debe obtener información del tiempo (en minutos) necesario para dar servicio a sus clientes. Para ello decidió seleccionar aleatoriamente 60 tiempos de servicio durante los meses de junio, julio y agosto del año 2013. Los mismos se muestran a continuación:

0,2	0,2	0,3	0,3	0,3	0,4	0,4	0,4	0,5	0,5	0,6	0,6	0,6	0,6	0,7
0,7	0,8	0,8	0,8	0,8	0,9	0,9	1,0	1,0	1,1	1,1	1,1	1,1	1,1	1,1
1,1	1,2	1,2	1,2	1,3	1,3	1,3	1,4	1,4	1,6	1,6	1,7	1,7	1,8	1,8
1,8	1,8	1,9	1,9	2,1	2,2	2,3	2,5	2,8	3,1	3,1	3,1	3,6	4,5	5,2

- a) Construya una gráfica de tallo y hoja para los datos.
- b) ¿Qué porcentaje de los tiempos de servicio son menores o iguales a 1 minuto?
- c) ¿Qué asimetría presentan estos tiempos de servicio? Interprete.
- d) Calcule $\{Q_1 - 1,5 (Q_3 - Q_1)\}$ y $\{Q_3 + 1,5 (Q_3 - Q_1)\}$. ¿A qué conclusión llega sobre los tiempos de servicio?
- e) Teniendo en cuenta las respuestas a los incisos anteriores, ¿hay tiempos atípicos? Si su respuesta es afirmativa y suponiendo que los mismos sean observaciones válidas, es decir que no hubo error en su recolección, ¿cómo explica su presencia a la Administración Central de esta cadena de supermercados?
- f) Si con estos datos construyera un histograma (no lo construya), ¿este gráfico mostraría la misma información que la gráfica hecha en el inciso a)?

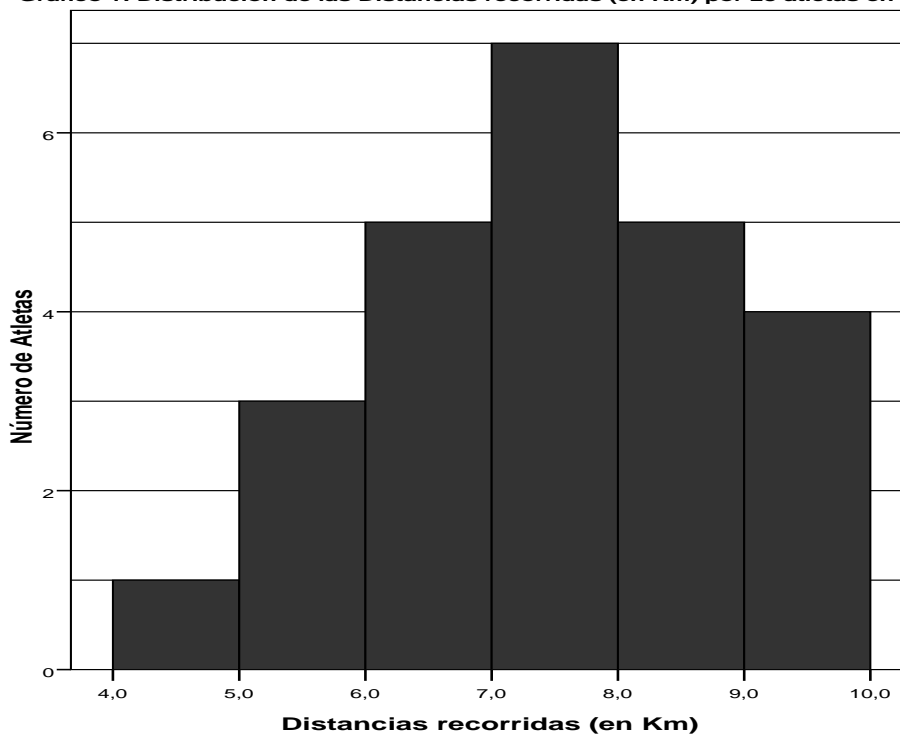
13) Los siguientes resultados corresponden a la distribución de frecuencias de una muestra de 25 distancias (en Km) recorridas por deportistas de un mismo sexo en un día de entrenamiento. El entrenador lleva este control porque espera que los 25 atletas se entrenen en forma similar. Él sabe que ese día el 75 % de los mismos recorrieron más de 6,5 km.

Medidas descriptivas de la muestra de distancias recorridas

Distancias recorridas (en km)	Medidas descriptivas	Estadístico
	Media	7,292
	Mediana	7,5
	Varianza	1,764
	Desvío típico	1,328
	Mínimo	4,6
	Máximo	9,5
	Rango	4,9
	Rango intercuartílico	1,7

El siguiente Histograma representa a la distribución de las distancias recorridas por los 25 atletas en un día de entrenamiento:

Gráfico 1: Distribución de las Distancias recorridas (en Km) por 25 atletas en un día de entrenamiento.



- a) ¿Habrá valores alejados y/o muy alejados en la muestra?

- b) Utilice alguna medida descriptiva para orientar al entrenador sobre la supuesta homogeneidad del trabajo de sus atletas.
- c) ¿Qué información está faltando en este gráfico?
- d) ¿Se ha utilizado la Regla de Sturges para construir el histograma?
- e) ¿Qué porcentaje de Atletas se encontrará por debajo de la media de distancias recorridas menos un desvío estándar?

14) Responda a las siguientes justificando.

- a) ¿Qué recomendaciones cree que debe tener en cuenta para la construcción de un diagrama de línea?
- b) De un diagrama de tallo y hoja, ¿se obtiene la misma información acerca del comportamiento frecuencialista de los datos, que de un histograma?
- c) ¿Puede una variable cualitativa tener una distribución con una asimetría a la derecha? Si su respuesta es afirmativa, proponga un ejemplo. Si su respuesta es negativa, justifique debidamente su respuesta.
- d) Cuando uno se refiere a una distribución simétrica, ¿esto hace alusión a una la distribución acampanada?
- e) ¿Es cierto que si los 50 datos numéricos de un lote están formados por un solo dígito no puedo mostrarlo a través de un diagrama de tallo y hoja?

15) Los empleados de una empresa usan para dirigirse a su trabajo diariamente el transporte público o el automóvil. A continuación aparecen los tiempos (en minutos) que 20 de ellos creen utilizar:

Transporte público	28	29	32	37	33	25	29	32	41	34
Automóvil	29	31	33	32	34	30	31	32	35	33

- a) Calcule las medidas descriptivas que crea convenientes para comparar estos lotes.
- b) En base a las medidas descriptivas calculadas en a), a su criterio, ¿cuál es el mejor medio de transporte? Justifique
- c) Utilice gráficos de caja para comparar ambas distribuciones. Interprete. ¿Sus conclusiones coinciden con las del inciso b)? Justifique.

16) En el período Noviembre-Diciembre 2013 se presentaron 45 aspirantes a un empleo. El criterio de selección es muy simple: "Se seleccionarán los aspirantes que logren tiempos de tarea que se encuentren hasta el primer decil de la distribución de frecuencias de los tiempos". A continuación se presentan los tiempos (en minutos) utilizados por cada aspirante al realizar dicha tarea.

Tiempos de Tarea de 45 aspirantes a un empleo

7,7	9,0	14,9	9,8	6,9	9,8	10,7	12,4	9,5	10,8	11,5	12,4	12,4	12,3	13,1
11,5	14,5	13,8	8,4	16,9	11,9	8,2	12,9	12,2	12,9	12,2	9,7	10,2	9,6	10,3
13,3	11,8	7,5	14,4	9,4	13,4	9,6	12,9	10,9	10,6	13,8	8,1	10,7	12,5	12,8

- a) Construya una distribución de frecuencias de la muestra de tiempos de tarea, agrupando los datos en intervalos de clase y grafique convenientemente.
- b) Analice la simetría de la distribución y realice un comentario sobre el comportamiento de los tiempos de tarea.
- c) Calcule la desviación estándar de los tiempos de tarea, ¿En qué unidades está expresada?
- d) Calcule el coeficiente de variación, ¿En qué unidades está expresado?
- e) ¿Qué conclusión puede extraer de lo calculado en los ítems c) y d).
- f) ¿Qué tiempos utilizaron los aspirantes que fueron seleccionados?

17) La compañía aérea “Veloz” desea mejorar el nivel de satisfacción de sus clientes. Para ello encuesta aleatoriamente a 19 usuarios durante el mes de junio de 2013 y les pide que expresen su grado de insatisfacción dando un valor numérico de 1 a 100. Un valor alto indica gran insatisfacción. Los resultados a continuación muestran un resumen estadístico de los niveles de insatisfacción dadas por esos clientes.

Grado de insatisfacción de clientes de la línea aérea “Veloz” – Junio 2013-

Profundidad	Tallo	Hoja				
1	3 _i	3				
1	3 _s					
6	4 _i	1 1 2 4 4				
(5)	4 _s	6 6 6 6 8				
8	5 _i	0 0 1 3 3 4				
2	5 _s	5				
1	6 _i	0	U = 1			

Fuente: Dpto. atención al cliente de la empres “Veloz”

$$\bar{X} = 47,53$$

$$M_e = 46$$

$$s_{n-1} = 6,23$$

- a) Construya un histograma de la distribución de frecuencias empíricas.
 - b) Calcule s_n utilizando el valor de s_{n-1} . Verifique el valor encontrado usando los datos.
 - c) Analice gráfica y analíticamente la forma de dicha distribución. Ambas respuestas, ¿son consistentes?
-

Respuestas TP1 – 2014 – fceUCSF -		
Ejercicio	Inciso	Respuesta
5	a	Aproximadamente 68%
	b	Aproximadamente 2,5
	c	Aproximadamente 84%
	d	Aproximadamente 13,5%
6	b	Si
7	a	No
	b	Si
	c	No
	d	No
8	Método A	
9	c	Media = 84,3 dólares Mediana = 75,5 dólares Cuartiles: 68 y 97,5 dólares Rango = 144 dólares Coeficiente de variación = 31,65%
	e	86,67%
11	a	Entre \$1875 y \$2777,78
	b	43 empleados
	c	Aproximadamente 16 empleados
12	b	40%
	c	As(+)
	d	$\{Q_1 - 1,5 (Q_3 - Q_1)\} = -0,95$ $\{Q_3 + 1,5 (Q_3 - Q_1)\} = 3,45$
	e	Si
	a	No
13	d	No
	e	15,568%
	b	Auto
16	c	$\bar{x} S_n = 2,1976$ minutos. Unidad de medida: minutos
	d	19,32%. Unidad de medida: no tiene
	f	Tiempos inferiores a 8,175 minutos (se usó la D^n de a)
17	b	$\bar{x} S_n = 6,06$