

Unidad Temática 1

Estadística descriptiva y análisis exploratorio de datos

1.1. Aplicación al caso del DERRAME MEDIO ANUAL del Río Mendoza

El derrame anual de un río es el volumen de agua registrado a lo largo de un año en un lugar determinado. La siguiente información corresponde al análisis estadístico realizado por los técnicos de la estación de mediciones para el derrame medio anual del Río Mendoza en Cacheuta, en hectómetros cúbicos, registrado durante 81 años consecutivos. Tenga presente que por cada año medido se ha obtenido su correspondiente derrame medio. Para elaborar un informe a partir de los resultados del Cuadro 1, Usted debería responder previamente las preguntas del siguiente cuestionario, con la justificación correspondiente:

- a) Interprete el valor de la mediana del derrame medio anual, sin omitir la unidad de medida.
- b) Describa brevemente la forma de la distribución de frecuencias de los derrames medios anuales del río, a partir del histograma de frecuencias.
- c) ¿Qué proporción de derrames medios anuales iguala o supera los 2211 hm³?
- d) ¿Qué porcentaje de derrames medios anuales son iguales o inferiores a los 3000 hm³?
- e) ¿Qué proporción de años el río entregó derrames medios anuales comprendidos entre 2000 y 2500 hm³?
- f) De acuerdo al comportamiento de los derrames medios anuales del río mostrado en el gráfico de caja, ¿cómo clasificaría los derrames medios anuales mayores de 2472,5 hm³. Justifique su respuesta.
- g) ¿En cuántos años del período registrado el río entregó un derrame medio anual igual o menor que 2000 hm³?
- h) ¿Entre qué valores se encuentra comprendida la mitad central de los derrames medios anuales registrados, ordenados de manera creciente?
- i) ¿En cuántos años del período registrado el río entregó un derrame anual superior a los 2500 hm³?
- j) Interprete el valor de la desviación estándar del derrame medio anual.
- k) Identifique el percentil 90 en una ojiva.

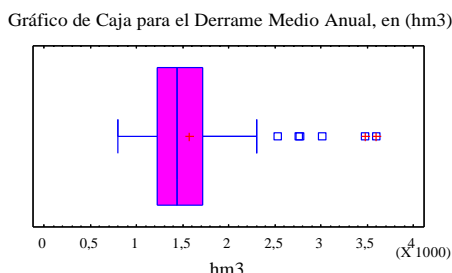
Cuadro 1

Resumen de las estadísticas	Percentiles
Número de observaciones = 81	
Media aritmética = 1578,75	
Mediana = 1440,0	P1 = 803
Moda =	P5 = 939
Varianza = 297878,0	P10 = 1081
Desviación Estándar = 545,782	P25 = 1225
Mínimo = 803,0	P50 = 1440
Máximo = 3597,0	P75 = 1724
Rango = 2794,0	P90 = 2211
Cuartil inferior = 1225,0	P95 = 2754
Cuartil superior = 1724,0	P99 = 3597
Rango intercuartílico = 499,0	
Coefficiente de variación = 34,5704%	

Tabla de frecuencias

Clase	Límites de Clase		Marca Clase	Frecuencias Simples		Frecuencias Acumuladas	
	Inferior	Superior		Absolutas	Relativas	Absolutas	Relativas
1	(500	1000]	750	5	0,0617	5	0,0617
2	(1000	1500]	1250	40	0,4938	45	0,5556
3	(1500	2000]	1750	23	0,2840	68	0,8395
4	(2000	2500]	2250	7	0,0864	75	0,9259
5	(2500	3000]	2750	3	0,0370	78	0,9630
6	(3000	3500]	3250	2	0,0247	80	0,9877
7	(3500	4000]	3750	1	0,0123	81	1,0000

Gráfico 1. Representación del Derrame Medio Anual del Río Mendoza en Cacheuta.



1.2. Aplicación al CONTROL DE CALIDAD DEL HORMIGÓN

Las especificaciones del pliego de condiciones indican que el hormigón a utilizar en la estructura de un edificio debe tener una resistencia característica de 17 MPa. Tenga en cuenta que resistencia característica es aquella resistencia que es superada por el 95% del total de resultados de ensayos a compresión disponibles a la edad de 28 días. A partir de la siguiente tabla de frecuencias construida con los resultados de ensayo de una muestra de 53 pastones de hormigón para la resistencia a compresión del hormigón a la edad de 28 días:

Cuadro 8. Tabla de frecuencias para la resistencia a compresión del hormigón a la edad de 28 días.

Clase i	Límites de Clase		Marca de Clase	FRECUENCIAS			
	(Linf	Lsup]		ABSOLUTAS		RELATIVAS	
				fi	Fi	fri	Fri
1	16,5	17,5			2		
2	17,5	18,5		6			
3	18,5	19,5			20		
4	19,5	20,5		19			
5	20,5	21,5			48		
6	21,5	22,5			52		
7	22,5	23,5					

- Construya una ojiva para la resistencia a compresión a la edad de 28 días.
- Del gráfico, obtenga el valor numérico aproximado y marque la resistencia característica del hormigón.
- El hormigón ensayado, ¿cumple las especificaciones del pliego? Justifique su respuesta.

1.3. El caso de las bolsitas de té

Una característica de calidad interesante en el proceso de llenado de bolsas de té es el peso de cada una. Si las bolsas tienen menos de lo que deben, surgen dos problemas. Primero, es posible que los clientes no puedan preparar un té tan fuerte como lo deseen. Segundo, la empresa violará las leyes de defensa al consumidor por la veracidad de las etiquetas. En este ejemplo, la etiqueta del paquete indica que, en promedio hay 5,5 gramos de té por bolsa. Por otro lado, si la cantidad promedio de té en la bolsa excede el peso indicado, la compañía regala parte del producto. Llenar las bolsas de té con la cantidad exacta de té es difícil debido a las variaciones de temperatura y humedad en el interior de la fábrica, las diferencias en la densidad del té y la alta velocidad con que trabaja la máquina de llenado (cerca de 170 bolsas por minuto).

Cuadro 9.1. Estadística descriptiva de la información registrada de la producción de una máquina durante una hora:

Estadísticas Descriptivas	Percentiles	Gráfico de Tallos y Hojas. Léase 54 7 = 5,47 gramos
Tamaño de la muestra = 50 Promedio = 5,5014 Mediana = 5,515 Moda = 5,53 Varianza = 0,0112 Desviación Estándar = 0,10583 Valor mínimo = 5,25 Valor máximo = 5,77 Rango = 0,52	1,0% = 5,25 5,0% = 5,32 10,0% = 5,35 25,0% = 5,44 50,0% = 5,515 75,0% = 5,57 90,0% = 5,625 95,0% = 5,67 99,0% = 5,77	2 52 59 5 53 224 6 53 6 15 54 000122444 21 54 556779 (11) 55 0001233344 18 55 5566777888 8 56 1123 4 56 577 HI 5,77

Cuadro 9.2. Tabla de Frecuencias

Clase	Límites de Clase		Punto Medio	Frecuencias Simples		Frecuencias Acumuladas	
	Inferior	Superior		Absoluta	Relativa	Absolutas	Relativas
1	[5,24	5,32]	5,28	4	0,0800	4	0,0800
2	(5,32	5,40]	5,36	5	0,1000	9	0,1800
3	(5,40	5,48]	5,44	11	0,2200	20	0,4000
4	(5,48	5,56]	5,52	16	0,3200	36	0,7200
5	(5,56	5,64]	5,60	10	0,2000	46	0,9200
6	(5,64	5,72]	5,68	3	0,0600	49	0,9800
7	(5,72	5,80]	5,76	1	0,0200	50	1,0000

Resolver

- En el contexto del problema, interprete el valor de la mediana
- En el contexto del problema, interprete el valor de la desviación estándar
- De acuerdo a la información disponible en la Tabla de Frecuencias, la Clase 6 contiene tres observaciones. ¿Cuáles son los valores numéricos de estas tres observaciones? Utilice la información disponible en el Anexo para responder esta pregunta. Justifique.
- ¿Qué *porcentaje* de bolsas de té contienen un peso superior a 5,64 gramos? Justifique.
- ¿Qué *proporción* de bolsas de té contienen un peso mayor de 5,4 gramos y no mayor que 5,64 gramos? Justifique.
- ¿Qué *cantidad* de bolsas de té contienen un peso no mayor de 5,48 gramos? Justifique.
- Indique en el gráfico de caja y extensión los valores utilizados para su construcción. ¿Por qué aparece en el gráfico un cuadradito en el extremo derecho? Justifique.
- ¿Es posible identificar una o más clases modales? Si es posible, indique cuál o cuáles son.
- En el contexto del problema, interprete el valor numérico del percentil noventa:
- El 80% de las bolsas de té pesa gramos o menos y el 20% de las bolsas de té pesa gramos o más. Utilice la representación gráfica disponible más adecuada para responder esta pregunta y marque en el gráfico su respuesta.
- Construya el gráfico de caja y extensión para el peso de las bolsas de té.
- Construya el histograma de frecuencias para el peso de las bolsas de té.
- Construya la ojiva para el peso de las bolsas de té.

1.4. Para responder usando el pensamiento y el razonamiento estadístico

¡Responder sin utilizar calculadora!

Suponga que Usted está investigando el tiempo de ensamblado de una pieza y para ello ha medido el tiempo de ensamblado de la pieza registrado por doce técnicos, en minutos. **Proponga** para cada una de las siguientes consignas, un conjunto de doce mediciones, en minutos. (Son doce mediciones para cada ítem). **Justifique**.

- El promedio del tiempo de ensamblado de las piezas debe ser menor que la mediana.
- La distribución del tiempo debe ser sesgada a la izquierda.
- El percentil setenta y cinco del tiempo de ensamblado debe ser igual al segundo cuartil.
- Si se construyera un gráfico de caja y extensión con las mediciones, la caja no debe tener extensiones.

1.5. El caso del método de capacitación

En un programa de capacitación industrial, un grupo de aprendices es instruido con el método A y otro grupo con el método B. En el método A, durante cinco días los aprendices dejan de trabajar en la planta industrial y asisten a un curso de capacitación que dura cuarenta horas. En el método B, los aprendices son capacitados en la misma planta industrial bajo la supervisión de expertos de la planta durante el mismo tiempo. Al finalizar la capacitación son evaluados y calificados. Diez aprendices son seleccionados aleatoriamente después de haber sido capacitados por cada uno de los métodos, son evaluados y calificados. Los datos de las calificaciones obtenidas por los aprendices son los siguientes:

Cuadro 1: Calificaciones

Método A:	71	75	65	69	73	66	68	71	74	68
Método B:	72	77	84	78	69	70	77	73	65	75

Cuadro 2: Estadística descriptiva de las calificaciones

Método A	Método B	Diagrama de tallos y hojas para Método A	Diagrama de tallos y hojas para Método B
Media = 70	Media = 74	6 56889	6 59
Mediana = 70	Mediana = 74	7 1134	7 023
Modas = 68 y 71	Moda = 77	7 5	7 5789
Desv. estándar = 3,37	Desv. estándar = 3,40		8 4
Calificación mínima = 65	Calificación mínima = 65	Percentiles Método A	Percentiles Método B
Calificación máxima = 75	Calificación máxima = 84	Percentil 05 = 65	Percentil 05 = 65
Rango = 10	Rango = 19	Percentil 10 = 65,5	Percentil 10 = 67
Primer cuartil = 68	Primer cuartil = 70	Percentil 70 = 72	Percentil 70 = 77
Tercer cuartil = 73	Tercer cuartil = 77	Percentil 90 = 74,5	Percentil 90 = 81
Rango intercuartil = 5	Rango intercuartil = 7	Percentil 95 = 75	Percentil 95 = 84
Coef. de variación = 4,8%	Coef. de variación = 7,3%		

Resolver: Al contestar cada uno de los siguientes ítems, tenga en cuenta que sólo una de las primeras cuatro opciones de cada uno de ellos es la correcta. Seleccione la opción correcta encerrando en un círculo la letra que la identifica. Debería saber justificar su respuesta.

- (1) Suponga que NO tiene los datos del Cuadro N° 1 y sólo dispone de la información del Cuadro N° 2. Si la **mediana** de las calificaciones de los aprendices capacitados con el método A es de 70 puntos, entonces se debe interpretar que:
 - a) La mitad de los aprendices obtuvo menos de 70 puntos y la otra mitad obtuvo más de 70.
 - b) Hubo 70 aprendices que fueron evaluados y obtuvieron un puntaje del 50%.
 - c) La mitad de los aprendices obtuvo 70 puntos o menos y la otra mitad obtuvo 70 puntos.
 - d) es/son correcta/s.
 (Completar con: *Todas las anteriores* o *Ninguna de las anteriores* o *Sólo las opciones ... y ...son correctas*)
- (2) De acuerdo a las medidas de dispersión, se pueden realizar las siguientes observaciones:
 - a) El 7,3% de los aprendices capacitados con el método B obtuvo una calificación inferior a la media (74 puntos) y otro 7,3% obtuvo una calificación superior a la media.
 - b) Para el método A, la mitad de las calificaciones están a 3,37 puntos a la izquierda de la media (74 puntos) y la otra mitad está a 3,37 puntos a la derecha de la media.
 - c) Las calificaciones obtenidas con el método B están más dispersas que las obtenidas con el método A.
 - d) es/son correcta/s.
 (Completar con: *Todas las anteriores* o *Ninguna de las anteriores* o *Sólo las opciones ... y ...son correctas*)
- (3) Teniendo en cuenta la información del Cuadro N° 1 y la del Cuadro N° 2, se puede afirmar que:
 - a) El diagrama de tallos y hojas del método B es correcto y el del método A es incorrecto.
 - b) El diagrama de tallos y hojas del método B es incorrecto y el del método A es correcto.
 - c) Ambos diagramas de tallos y hojas son correctos.
 - d) Ambos diagramas de tallos y hojas son incorrectos.
 - e) es/son correcta/s.
 (Completar con: *Todas las anteriores* o *Ninguna de las anteriores* o *Sólo las opciones ... y ...son correctas*)
- (4) Si se representara un **gráfico de caja y extensión** para las calificaciones obtenidas por el método A y otro para las calificaciones obtenidas por el método B, se pueden realizar las siguientes afirmaciones:
 - a) El diagrama de caja correspondiente al método A sería simétrico.

- b) Si se representa un diagrama debajo del otro, se produciría una superposición de las cajas, al menos en parte de ellas, en la escala de las calificaciones.
- c) El diagrama de caja correspondiente al método B presentaría un dato apartado.
- d) es/son correcta/s.
(Completar con: *Todas las anteriores* o *Ninguna de las anteriores* o *Sólo las opciones ... y ...son correctas*)
- (5) Si la evaluación se aprueba con un **mínimo de 70 puntos**, a partir de la información del Cuadro N° 2, se puede afirmar que:
- a) El 72% de los aprendices capacitados con el método A, aprobará la evaluación.
- b) El 77% de los aprendices capacitados con el método B, aprobará la evaluación.
- c) El 70% de los aprendices capacitados con el método A aprobará la evaluación con un puntaje igual a 72, mientras que el 70% de los aprendices capacitados con el método B aprobará la evaluación con un puntaje igual a 77.
- d) es/son correcta/s.
(Completar con: *Todas las anteriores* o *Ninguna de las anteriores* o *Sólo las opciones ... y ...son correctas*)

1.6. Influencia del tipo de neumáticos en el rendimiento del combustible

Se lleva a cabo un estudio para analizar la influencia del uso de neumáticos radiales en lugar de neumáticos comunes en la economía de combustible. Se equipan 50 autos con neumáticos radiales y luego, los mismos autos, se equipan con neumáticos comunes y se manejen por un recorrido de prueba.

El consumo de combustible, en kilómetros por litro, se registró en el cuadro siguiente:

Cuadro. Consumo de combustible, en kilómetros por litro.

#	Radiales	Comunes	#	Radiales	Comunes	#	Radiales	Comunes	#	Radiales	Comunes	#	Radiales	Comunes
1	4,2	4,1	11	6,1	6,0	21	7,5	7,0	31	5,6	5,6	41	6,9	7,0
2	4,7	4,9	12	5,2	4,9	22	5,0	4,8	32	5,9	5,7	42	6,8	6,7
3	6,6	6,2	13	4,3	4,2	23	6,2	6,1	33	7,3	6,8	43	4,4	4,5
4	7,0	6,9	14	4,8	5,0	24	5,3	5,0	34	4,8	4,6	44	5,8	5,6
5	6,7	6,8	15	6,7	6,3	25	4,1	4,0	35	6,0	5,9	45	5,9	5,9
6	4,5	4,4	16	7,1	7,0	26	4,6	4,8	36	5,1	4,8	46	7,5	6,8
7	5,7	5,7	17	6,8	6,9	27	6,5	6,1	37	4,6	4,8	47	4,8	4,8
8	6,0	5,8	18	4,6	4,5	28	6,9	6,8	38	4,3	4,0	48	6,2	5,9
9	7,4	6,9	19	5,8	5,8	29	6,6	6,7	39	4,6	5,0	49	5,1	5,0
10	4,9	4,7	20	6,1	5,9	30	4,4	4,3	40	6,7	6,1	50	4,3	4,0

- a) Calcule e interprete las medidas de tendencia central: media, mediana y moda.
- b) Calcule e interprete las medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación.
- c) Describa el patrón del comportamiento de los datos de ambos tipos de neumáticos.
- d) Establezca comparaciones entre los neumáticos, a partir del análisis exploratorio de datos.

1.7. Influencia de la temperatura en la duración de las bombas de combustible

Los siguientes datos representan la duración de la vida, en años, de 30 bombas de combustibles similares, a temperaturas extremas:

Cuadro. Duración, en años, de 30 bombas de combustibles.

#	Cálido	Frío	#	Cálido	Frío	#	Cálido	Frío	#	Cálido	Frío	#	Cálido	Frío	#	Cálido	Frío
1	2,0	1,8	6	3,0	2,9	11	0,3	0,2	16	3,3	3,2	21	1,3	1,1	26	0,4	0,3
2	0,2	0,1	7	6,0	5,8	12	5,5	5,4	17	6,5	6,3	22	0,2	0,3	27	2,3	2,1
3	1,5	1,3	8	4,0	4,1	13	5,9	5,7	18	1,8	1,9	23	4,7	4,5	28	0,7	0,6
4	4,5	4,4	9	0,3	0,1	14	1,5	1,4	19	0,5	0,4	24	2,5	2,4	29	5,0	4,8
5	1,0	0,8	10	6,0	5,9	15	5,6	5,4	20	6,0	5,9	25	1,2	1,0	30	0,2	0,4

- a) Calcule e interprete las medidas de tendencia central: media, mediana y moda.
- b) Calcule e interprete las medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación.
- c) Describa el patrón del comportamiento de los datos de ambos tipos de neumáticos.
- d) Establezca comparaciones entre los neumáticos, a partir del análisis exploratorio de datos.

1.8. Aplicación al cálculo de UNIONES DE MADERA ATORNILLADAS

Se ensayaron 15 uniones de maderas atornilladas con bullones de 12 milímetros de diámetro sometidas a esfuerzo cortante doble. Las uniones estaban formadas por tres piezas de pino Paraná de sección nominal (2"x4"). Los resultados del ensayo, en kg, se indican en el cuadro, donde se definen las siguientes variables:

X: carga que produce una deformación de 1,5 milímetros.

Y: carga que produce una deformación de 3 milímetros.

Z: carga última que es capaz de soportar la unión.

Cuadro. Resultados de ensayo de 15 uniones de madera atornilladas con bulones de 12 mm, en kg.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	1750	1700	1550	1350	1450	1700	1350	1550	1550	1140	1250	1600	1300	1700	1600
y	1950	1750	1800	1750	1840	2125	1725	1850	2050	1600	1650	2000	1700	1925	2100
z	2125	1950	2050	2000	2050	2250	2000	2100	2200	1850	1925	2200	1950	2100	2250

Resolver los siguientes ítems usando Excel

- a) Representar la nube de puntos para la carga que produce una deformación de 3 milímetros y la carga última.
- b) Ajustar una recta por el método de mínimos cuadrados.
- c) Calcular el coeficiente de correlación muestral y el coeficiente de determinación muestral. Analizar los resultados.
- d) En función de los resultados anteriores, este ensayo simple y no destructivo, ¿puede utilizarse para obtener una buena predicción de la capacidad última de la unión abollonada?
- e) Repetir el análisis para la carga que produce una deformación de 1,5 milímetros y la carga última.
- f) Comparar los resultados obtenidos en ambos casos.

1.9. El caso del Método de Pesada

En uno de los viajes que realizaron Norberto y Daniel, profesores de Química Analítica y de Probabilidad y Estadística, respectivamente, Norberto le pidió a Daniel que estudiara algunas experiencias propuestas por Skoog, West, Holler y Crouch en su texto de Química Analítica, experiencias que tienen que ver con ambos espacios curriculares.

Daniel aceptó la propuesta de Norberto y trabajó con Daniela Fernández. Daniela, es una ingeniera química que realizaba una pasantía de articulación entre los espacios curriculares de Probabilidad y Estadística y Química Analítica.

Los autores del texto mencionado proponen una serie de experimentos, con el propósito de presentar varias de las herramientas, técnicas y habilidades necesarias para trabajar en un laboratorio de química analítica. Ellos describen cada una de las técnicas por separado y las operaciones unitarias. Consideran que para los estudiantes de química es muy importante aprender algunas técnicas estadísticas y adquirir las habilidades pertinentes antes de realizar otros experimentos de laboratorio.

El primero de los experimentos tiene que ver con el **manejo de la balanza analítica**. En síntesis, el experimento consiste en *comparar dos métodos de pesada*. En primer lugar se debe obtener la masa de un número dado de monedas nuevas, pesando cada una de ellas individualmente. Al resultado obtenido los denominaremos *Peso Individual*. A continuación, se debe determinar la masa de todas las monedas juntas, para después ir quitando una por una las monedas y calcular la masa individual de las mismas por diferencia. El resultado obtenido por este método lo denominaremos *Peso Por Diferencia*.

1. Objetivo

Utilizar herramientas del análisis estadístico para determinar el método de pesada más adecuado.

2. Procedimiento en el laboratorio de química

1. Seleccionar una muestra de 35 monedas de diez centavos (con muy poco o sin desgaste).

2. Posicionar cada una de las monedas en un papel rotulado que identifique la moneda a pesar.
3. Pesar, una a una, todas las monedas y registrar la masa de cada una, en gramos. (*Peso Individual*).
4. Colocar las mismas monedas en la balaza, todas juntas y sin perder la identificación de cada una, y determinar la masa total.
5. Retirar una moneda y determinar la masa de las que quedan en el platillo.
6. Por diferencia, establecer el peso la moneda que se retira. (*Peso Por Diferencia*).
7. Repetir esta operación apartando una moneda cada vez.

3. La experiencia de Daniela

Daniela pesó 35 monedas de diez centavos en la balanza del laboratorio por un método de pesada y el otro. La identificación de cada moneda es la que se presenta en el Cuadro 1. Los resultados de las masas individuales de las monedas que obtuvo, se presentan en los Cuadros 2 y 3.

Cuadro 1: Correspondencia ente la identificación de la moneda y los valores presentados en Cuadros 2 y 3.

01	-	02	-	03	-	04	-	05	-	06	-	07	-	08	-	09	-	10
11	-	12	-	13	-	14	-	15	-	16	-	17	-	18	-	19	-	20
21	-	22	-	23	-	24	-	25	-	26	-	27	-	28	-	29	-	30
31	-	32	-	33	-	34	-	35										

Cuadro 2: *Peso Individual* de las monedas de diez centavos, en gramos. Método Directo.

2,2441	-	2,2381	-	2,2036	-	2,2615	-	2,2664	-	2,1500	-	2,2270	-	2,3146	-	2,1869	-	2,2837
2,2333	-	2,2375	-	2,2554	-	2,2367	-	2,2453	-	2,1843	-	2,2581	-	2,1552	-	2,2054	-	2,2304
2,2446	-	2,2109	-	2,2573	-	2,2390	-	2,2438	-	2,2047	-	2,2755	-	2,1936	-	2,2416	-	2,2494
2,2256	-	2,2388	-	2,2547	-	2,1745	-	2,2326										

Cuadro 3: *Peso Por Diferencia* de las monedas de diez centavos, en gramos. Método por Diferencia de Pesadas.

2,2447	-	2,2397	-	2,2053	-	2,2645	-	2,2668	-	2,1496	-	2,2283	-	2,3174	-	2,1875	-	2,2861
2,2340	-	2,2384	-	2,2559	-	2,2405	-	2,2474	-	2,1852	-	2,2593	-	2,1546	-	2,2070	-	2,2304
2,2477	-	2,2106	-	2,2575	-	2,2391	-	2,2441	-	2,2041	-	2,2772	-	2,1939	-	2,2414	-	2,2504
2,2257	-	2,2391	-	2,2552	-	2,1781	-	2,2327										

4. Análisis estadístico de los pesos individuales de las monedas

4.1. Distribución de frecuencias

Daniela aplicó los procedimientos del análisis estadístico para construir la distribución de frecuencias con las masas individuales de las monedas obtenidas por ambos métodos de pesada y tabuló los resultados que obtuvo, construyendo así los Cuadros 4 y 5.

Cuadro 4: Tabla de frecuencias para la variable *Peso Individual*. Método Directo.

Clase	Límites de Clase (Inferior Superior]		Punto Medio	Frecuencias Absoluta	Simples Relativa	Frecuencias Absoluta	Acumuladas Relativa (prop.)	Acumuladas Relativa (%)
1	(2,13	2,16]	2,145	2	0,0571	2	0,0571	5,71%
2	(2,16	2,19]	2,175	3	0,0857	5	0,1429	14,29%
3	(2,19	2,22]	2,205	5	0,1429	10	0,2857	28,57%
4	(2,22	2,25]	2,235	16	0,4571	26	0,7429	74,29%
5	(2,25	2,28]	2,265	7	0,2000	33	0,9429	94,29%
6	(2,28	2,31]	2,295	1	0,0286	34	0,9714	97,14%
7	(2,31	2,34]	2,325	1	0,0286	35	1,0000	100,00%

Cuadro 5: Tabla de frecuencias para la variable *Peso Por Diferencia*. Método por Diferencia de Pesadas.

Clase	Límites de Clase (Inferior Superior]		Punto Medio	Frecuencias Absoluta	Simples Relativa	Frecuencias Absoluta	Acumuladas Relativa (prop.)	Acumuladas Relativa (%)
1	(2,13	2,16]	2,145	2	0,0571	2	0,0571	5,71%
2	(2,16	2,19]	2,175	3	0,0857	5	0,1429	14,29%
3	(2,19	2,22]	2,205	5	0,1429	10	0,2857	28,57%
4	(2,22	2,25]	2,235	15	0,4286	25	0,7143	71,43%
5	(2,25	2,28]	2,265	8	0,2286	33	0,9429	94,29%
6	(2,28	2,31]	2,295	1	0,0286	34	0,9714	97,14%

7	(2,31	2,34]	2,325	1	0,0286	35	1,0000	100,00%
---	-------	-------	-------	---	--------	----	--------	---------

Se ha adoptado el criterio de tomar abierto el límite inferior de clase y cerrado el límite superior de la misma: (]. Recuerde la importancia de armar las clases de modo tal que sean *mutuamente excluyentes* y *completamente exhaustivas*. La primera propiedad se refiere a que el peso de una misma moneda no puede estar en más de una clase a la vez. La segunda propiedad, hace referencia a que entre todas las clases deben incluir a la totalidad de las monedas que se pesaron; ninguna moneda debe quedar fuera de las clases establecidas.

Al *punto medio* de clase también se lo suele denominar *marca de clase*. En los Cuadros 4 y 5 los valores de las columnas 6 y 8 están expresados en términos de *proporciones*. Los valores de la columna 9 se corresponden con los de la columna 8 expresados en *porcentaje*.

4.2. Estadística descriptiva de la masa individual de las monedas

Daniela corrió el software con los datos de ambas variables (*Peso Individual* y *Peso Por Diferencia*) y tabuló los resultados en el Cuadro 6.

Cuadro 6: Estadística descriptiva de la masa de las monedas de diez centavos, obtenida por ambos métodos de pesada.

Estadística	Método 1: <i>Peso Individual</i>	Método 2: <i>Peso por Diferencia</i>	Diagrama de tallo y hojas Unidad = 0,01 Ejemplo : 21 5 representa 2,15 gramos.
Cantidad de observaciones	35	35	Muestra 1: Peso Individual
Promedio	2,23155	2,23258	2 21 55
Mediana	2,2381	2,2391	3 21 7
Moda		2,2391	6 21 889
Varianza	0,00122079	0,00124882	10 22 0001
Desviación estándar	0,0349399	0,0353387	(10) 22 2233333333
Mínimo	2,15	2,1496	15 22 4444445555
Máximo	2,3146	2,3174	5 22 667
Rango	0,1646	0,1678	2 22 8
Cuartil Inferior	2,2054	2,207	1 23 1
Cuartil Superior	2,2547	2,2552	
Rango intercuartil	0,0493	0,0482	Muestra 2: Peso Por Diferencia
Coeficiente de variación	1,56572%	1,58286%	2 21 45
Percentiles:			3 21 7
P01	2,15	2,1496	6 21 889
P05	2,1552	2,1546	10 22 0001
P10	2,1843	2,1852	(9) 22 223333333
P90	2,2664	2,2668	16 22 4444445555
P95	2,2837	2,2861	5 22 667
P99	2,3146	2,3174	2 22 8
			1 23 1

5. Representaciones gráficas de las masas individuales de las monedas

En la página siguiente se presentan los gráficos en dos columnas. Los de la izquierda corresponden a la masa de las monedas obtenidas por el método del *Peso Individual*. Los de la columna derecha corresponden al método del *Peso Por Diferencia*.

Después de estudiar La Unidad 1, usted debe saber construir todos estos gráficos, con excepción del gráfico cuyo título es *Trazado de la densidad*, que lo estudiaremos en la Unidad 3.

6. Transformación de variables: *valor Z*

Le proponemos la siguiente actividad. Ingrese al Sitio Web de la cátedra, diríjase a la página *Documentos de Apoyo* y lea el documento *04 Valor Z*.

Ahora debe responder la Autoevaluación que le proponemos en el apartado 1.7 del Caso. Los ítems en los que se evalúa el concepto de *valor Z* se deben responder razonando la respuesta, sin necesidad de realizar cálculos

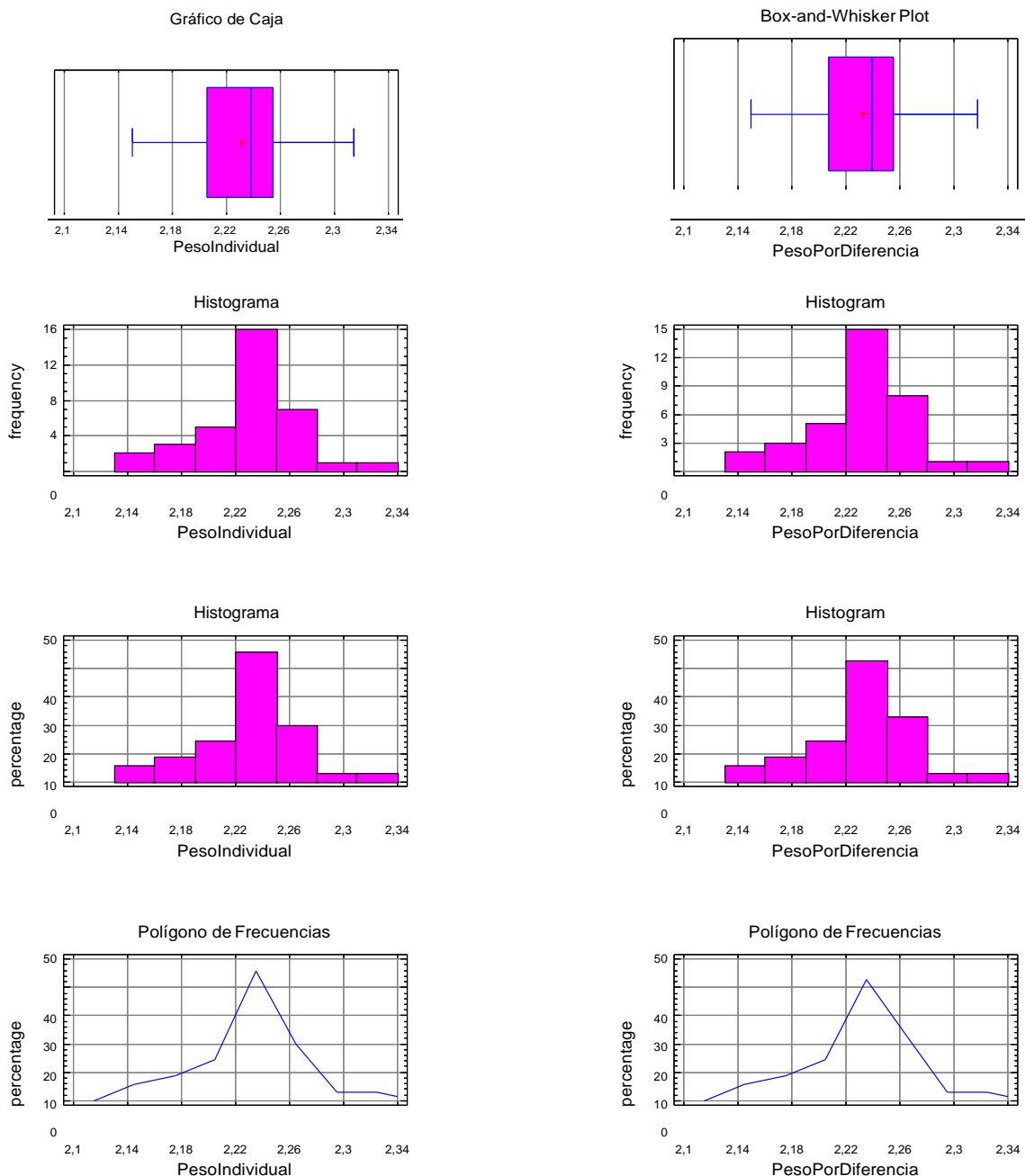
Una vez que haya completado la autoevaluación, vuelva a entrar en el Sitio Web de la cátedra, diríjase a la página *Base de Datos* y baje el archivo denominado *PesadaDeMonedas.xls*. Es un libro de Microsoft EXCEL que tiene dos Hojas. En la primera encontrará los *Datos* que Daniela registró como resultado de la experiencia de pesar las monedas en el laboratorio de química. En la segunda Hoja se han calculado los valores de *Z* de las observaciones obtenidas por uno y otro método de pesada.

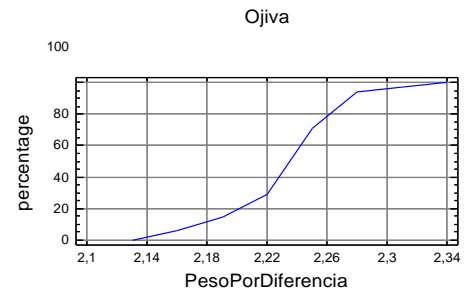
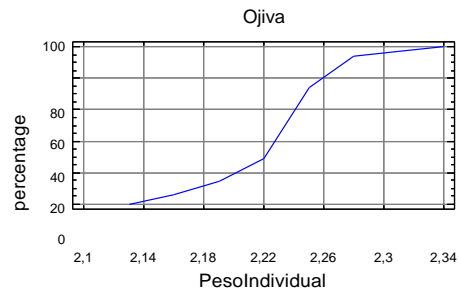
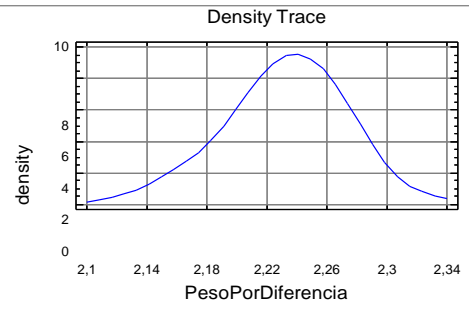
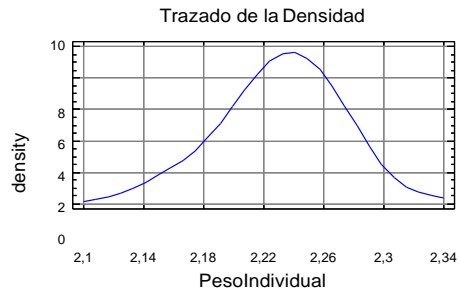
- Si el valor Z_X correspondiente al *PesoIndividual* de la Moneda N° 10 es igual a 1,5145, ¿cómo interpretaría este valor numérico?

En la Hoja denominada *valor Z* del Libro EXCEL modifique uno cualquiera de los valores de la variable X (*Peso Individual*), por ejemplo, cambie el valor del *Peso Individual* de la Moneda 30 al valor 2,0000 y observe que el valor de Z_X resultante es (−4,3531). Observe también que los valores de la media y desviación estándar de los valores de Z_X no cambian (*media* = 0 y *desviación estándar* = 1). Pruebe cualquier para el valor de x cualquier otro cambio y observe lo que sucede con el *valor z* y con la media y desviación estándar de z .

- Piense y luego proceda a realizar un cambio en el valor del *Peso Por Diferencia* de la Moneda N° 30 que conduzca a obtener un valor de Z_Y negativo. Propóngase no utilizar el método de “prueba y error”. Piense y luego decida el cambio propuesto para lograr lo que se piden en la consigna.

Gráfico 1: Representaciones gráficas para el caso del Método de Pesada, obtenidas de la exploración de los datos resultantes de pesar las monedas.





7. Autoevaluación

Al observar las representaciones gráficas nos deberíamos formular las siguientes preguntas y proponer las respuestas correspondientes.

1. ¿Para representar qué tipo de datos (categóricos, numéricos o ambos) se puede utilizar cada uno de los gráficos?
2. ¿En cuáles de los gráficos se puede observar la evidencia de simetría o sesgo de la distribución de frecuencias?
3. ¿Qué medidas descriptivas (tendencia central, dispersión, posición) se pueden leer:
 - a) En el gráfico de caja?
 - b) En el histograma?
 - c) En el polígono de frecuencias?
 - d) En la ojiva?
4. ¿Cómo cambiarían los gráficos de caja e histograma si la ojiva tuviese mayor o menor pendiente?
5. ¿Cómo se evidencia la concentración de datos en cada uno de los gráficos?
6. En el histograma, ¿qué ventaja tiene representar las frecuencias en términos relativos (en porcentaje, por ejemplo), en vez de hacerlo como frecuencias absolutas?

Después de estudiar la Unidad Temática 1, usted debe estar en condiciones de responder los siguientes ítems de la autoevaluación. Al contestar cada uno de ellos, tenga en cuenta que sólo una de las primeras cuatro opciones de cada ítem es la correcta. Seleccione la opción correcta encerrando en un círculo la letra que la identifica.

- a b c d (1) El número de clases en que se puede agrupar los datos individuales para construir la distribución de frecuencias:
- a) Para el tamaño de la información disponible, podría estimarse con la fórmula de Sturges o bien como la raíz cuadrada de n .
 - b) Se podría adoptar un número comprendido entre 5,92 y 6,10.
 - c) No es incorrecto adoptar 7 clases, si la exploración de los datos indica que este número de clases interpreta mejor el comportamiento de los datos.
 - d) Todas las anteriores.
- a b c d (2) Teniendo en cuenta sólo la información del Cuadro 6 para el **Peso Individual** se debe concluir que:
- a) Al menos un cuarto de las monedas pesan 2,2054 gramos o más.
 - b) La mitad de las monedas pesan 2,2381 gramos.
 - c) La suma de las *desviaciones* del peso de las monedas respecto de la media es 0,0349399.
 - d) Ninguna de las anteriores.
- a b c d (3) Teniendo en cuenta sólo la información del Cuadro 6 para el **Peso Por Diferencia**, se concluye:
- a) El promedio de las *desviaciones cuadráticas* respecto del peso promedio es 0,00124882.
 - b) Al menos la mitad de las monedas pesan 2,2391 gramos o más.
 - c) Al menos la mitad de las monedas pesan entre 2,207 y 2,2552 gramos.
 - d) Todas las anteriores.
- a b c d (4) Teniendo en cuenta sólo la información de la tabla de frecuencias para el **Peso Por Diferencia** de las monedas (Cuadro 5), se concluye que:
- a) Hay tres monedas cuyo peso excedió los 2,16 gramos pero no sobrepasó los 2,175 gramos.
 - b) El 5,71% de las monedas tuvieron un peso que no sobrepasó los 2,16 gramos.
 - c) El peso de diez monedas excedió los 2,22 gramos.
 - d) Todas las anteriores.
- a b c d (5) Teniendo en cuenta sólo la información disponible en el Cuadro 6 para el **Peso Por Diferencia**, se concluye:
- a) El peso de algunas monedas da lugar a la presencia de *datos apartados*.
 - b) El peso de sólo una de las monedas debe considerarse como *dato apartado*.
 - c) Los datos apartados serían fácilmente identificables en una *ojiva*.
 - d) Ninguna de las anteriores.

- a b c d (6) Teniendo en cuenta sólo la información del Cuadro 6 para el *Peso Por Diferencia*, se concluye que:
- a) El percentil 92 es menor de 2,2668 gramos.
 - b) El percentil 25 es igual a 2,2552 gramos.
 - c) Al menos el 5% del peso de las monedas de diez centavos excede de 2,1546 gramos.
 - d) El percentil 97 *podría* ser igual o mayor que 2,2861 gramos.
- a b c d (7) La información del Cuadro 6 ha sido obtenida corriendo el software con los datos registrados por Daniela. Observe que el programa ha reducido la información del diagrama de tallo y hojas a dos decimales. De la lectura del diagrama para el *Peso Individual*, se debe interpretar que:
- a) Seis monedas, pesaron 2,18 gramos o menos.
 - b) Cinco monedas, pesaron 2,26 gramos o más.
 - c) Dos monedas, pesaron 21,5 gramos.
 - d) Todas las anteriores.
- a b c d (8) Teniendo en cuenta sólo la información del Cuadro 6 para el *Peso Por Diferencia*, la masa de una moneda será considerada como un *dato apartado*:
- a) Si es inferior a 2,3275 gramos.
 - b) Si se encuentra entre 2,1347 y 2,3275 gramos.
 - c) Si es inferior a 2,1347 gramos.
 - d) Si es superior a 2,1347 gramos.
- a b c d (9) De la lectura del Cuadro 5, se debe interpretar que:
- a) Cinco monedas pesaron más de 2,25 gramos, pero no superaron los 2,28 gramos.
 - b) Ocho monedas pesaron 2,265 gramos.
 - c) El 5,71% de las monedas pesó más de 2,28 gramos.
 - d) 0,2286 es la proporción de monedas que registró un peso mayor de 2,25 gramos.
- a b c d (10) Para responder el ítem, tenga en cuenta el Cuadro 4 y la representación gráfica correspondiente al *Peso Individual*.
- a) En la distribución de frecuencias del *Peso Individual* se observa una única clase modal.
 - b) De la lectura de la ojiva del *Peso Individual* se observa que, aproximadamente, el 80% de las monedas pesaron 2,26 gramos o menos.
 - c) El polígono de frecuencias puede interpretarse como una “aproximación gruesa” del Trazado de la Densidad.
 - d) Todas las anteriores.
- a b c d (11) Podría describirse el *patrón de comportamiento* del *Peso Individual* de las monedas de diez centavos diciendo que:
- a) La masa de las monedas tiene una distribución aproximadamente simétrica.
 - b) Presenta una única clase modal; esto es, se observa una concentración de monedas que tienen una masa comprendida entre 2,22 y 2,25 gramos.
 - c) Ninguna moneda pesó menos de 2,13 gramos ni más de 2,34 gramos.
 - d) Todas las anteriores.
- a b c d (12) Si el *valor Z* de una de las monedas que pesó Daniela es igual a 1,5145, se debe interpretar que:
- a) El peso de dicha moneda es igual a 1,5145 gramos.
 - b) El peso de dicha moneda está por encima del peso promedio de las monedas.
 - c) El peso de la moneda está por debajo del peso promedio de las monedas.
 - d) Ninguna de las anteriores.

Respuestas para el caso del Método de Pesada.

1.d) – 2.d) – 3.d) – 4.b) – 5.d) – 6.d) – 7.d) – 8.c) – 9.c) – 10.d) – 11.d) – 12.b)