

Unidad Temática 1

Estadística descriptiva y análisis de datos

Responda verdadero o falso. Coloque una letra V a la izquierda del número del ítem si acepta la afirmación enunciada, o una F si la rechaza.

1. Definiciones preliminares, tipos de datos y variables

- 1. Desde el punto de vista estadístico, las *poblaciones* pueden ser *finitas* o *infinitas*.
- 2. Si se ha recopilado la información deseada para los elementos de un *subconjunto* que representa a una población objeto de estudio, se está en presencia de un *censo*.
- 3. Si el Ministro de Educación está interesado en el rendimiento de los estudiantes argentinos, medido por el promedio de calificaciones, las *unidad de análisis* son los establecimientos educativos del país.
- 4. La rama de la estadística que se ocupa de utilizar datos de una *muestra* para hacer inferencias acerca de la *población* en estudio, se conoce con el nombre de *estadística descriptiva*.
- 5. El nivel de estudios o escolaridad de los empleados de una empresa es una variable cuantitativa.
- 6. El número de hijos de los trabajadores de una fábrica es una variable cuantitativa continua.
- 7. El color de ojos de las personas es una variable cualitativa que se mide en una escala nominal.
- 8. Una escala nominal consiste en categorías mutuamente excluyentes que no implican un orden jerárquico entre ellas.
- 9. El cargo que ocupa un empleado en la empresa, es una variable cualitativa y se mide en escala ordinal.
- 10. La antigüedad de un empleado en una institución pública es una variable numérica que se mide en una escala nominal.
- 11. La escala de intervalo es una forma de medida más completa que la escala ordinal, ya que permite discernir no sólo qué valor observado es el más grande, sino también por cuánto.
- 12. Los *datos primarios* son siempre de mejor calidad que los *datos secundarios*.
- 13. Los *censos*, en general, resultan muy costosos, difíciles de realizar e incluso en algunos casos pueden resultar imposibles de llevar a cabo.
- 14. El número del piso desde el que es llamado un ascensor de un edificio en altura, es una variable *numérica continua*.
- 15. El *promedio* de los resultados obtenidos al lanzar dos dados, es una variable *numérica continua*.
- 16. Cuando el conjunto de valores que puede tomar una variable es *finito*, es decir, se puede *contar*, se dice que la variable es *discreta*.

- 17. Las variables *continuas* son aquellas en que los datos resultantes de las mediciones, pueden tomar cualquiera de los valores de una escala continua, en el rango para el cual está definida la variable. De otro modo, pueden tomar el continuo de valores entre el mínimo y el máximo observado.
- 18. Las gráficas circulares y gráficas de barras, son herramientas útiles para descripción gráfica de conjuntos de datos *cuantitativos*.
- 19. Las gráficas circulares también se las conoce con el nombre de gráficas de pastel, gráficas de torta o gráficas de sectores.
- 20. Las gráficas de barras pueden representarse tanto con barras verticales como horizontales.
- 21. En el diagrama de Pareto, las *categorías* de la variable cualitativa deben disponerse en orden decreciente por altura (o frecuencia) y se muestra una poligonal acumulativa superpuesta a las barras.
- 22. Para representar las gráficas de barras o las circulares, se pueden emplear tanto frecuencias absolutas como relativas.
- 23. La representaciones de las gráficas de barras y gráficas circulares en *escala relativa* (proporción o porcentaje), tienen la ventaja de independizarse del tamaño de la muestra a partir de la cual se obtuvo la información.
- 24. Las gráficas de sectores resultan más apropiadas, es decir, más cómodas de leer, cuando se tiene variables cualitativas con una gran cantidad de categorías, por ejemplo 26.
- 25. Cuando se tiene una variable *cualitativa*, las categorías en que se agrupan los datos para representarlos mediante una gráfica de barras, son mutuamente excluyentes.
- 26. En las gráficas de barras, estas últimas pueden ir pegadas una a la otra.
- 27. Si la población estudiantil de una Universidad es de 12.000 alumnos, para representar gráficamente el *turno* en que cursan los estudiantes (mañana, tarde o noche), se puede utilizar tanto una gráfica de sectores como una gráfica de barras.
- 28. La gráfica de puntos encuentra su mejor aplicación en el caso de conjuntos de datos pequeños.
- 29. Las *distribuciones de frecuencias* sacrifican algunos detalles, pero ofrecen información acerca del *patrón de comportamiento* de los datos.
- 30. El *histograma* es una representación gráfica que se utiliza para representar tanto variables *numéricas* como *no numéricas*.
- 31. Al agrupar los datos en tablas de frecuencias, un dato particular del conjunto de datos, debe pertenecer a una y sólo una clase o categoría, por lo que se dice que las clases son *completamente inclusivas*.
- 32. La organización de los datos por categorías o clases permite identificar *patrones de comportamientos* evidentes de los mismos.
- 33. A partir de una *distribución de frecuencias*, se puede reconstruir una lista con la totalidad de los datos observados a partir de la cual se construyó dicha tabla.

- 34. Una *distribución de frecuencias* es una tabla en la que organizamos los datos en *clases* y se muestra el número de observaciones del conjunto de datos que caen en cada una de las clases.
- 35. Las clases en que se agrupan los datos en una tabla de frecuencias, deben ser *completamente inclusivas*, esto significa que, los datos que más se repiten deben incluirse en una misma clase.
- 36. Las *clases* de una distribución de frecuencias deben ser mutuamente excluyentes y completamente inclusivas.
- 37. De ser posible, *conviene* que todas las clases de una distribución de frecuencias tengan el mismo ancho; de lo contrario, tendríamos una distribución mucho más difícil de interpretar.
- 38. Para construir las distribuciones de frecuencias, como regla general, los estadísticos aconsejan utilizar entre 20 y 25 clases.
- 39. El *número de clases* que se utiliza para construir la distribución de frecuencias es independiente de la cantidad de datos disponibles.
- 40. La fórmula de Sturges, da el número de clases de una distribución de frecuencias y puede utilizarse para iniciar la exploración a partir de la misma: $k = 1 + 3,3 \log n$.
- 41. Sea cual sea el conjunto de datos que se desea representar gráficamente, para determinar el número de clases, es indistinto emplear la fórmula de Sturges o la fórmula \sqrt{n} .
- 42. Las gráficas de distribuciones de frecuencias simples y de distribuciones de frecuencias relativas resaltan y aclaran los *patrones* que no se pueden distinguir fácilmente en las tablas.
- 43. Los *histogramas* se pueden construir utilizando tanto las frecuencias absolutas como las frecuencias relativas.
- 44. La frecuencia relativa simple de cualquier clase particular, se obtiene calculando el cociente entre el número de observaciones que entran en la clase y el número total de observaciones realizadas.
- 45. En una distribución de frecuencias, la suma de todas las *frecuencias relativas* de todas las clases, es igual al número total de observaciones realizadas.
- 46. Cuando el *histograma* se construye utilizando frecuencias relativas, resulta fácil comparar los datos de muestras de tamaños diferentes.
- 47. La *marca de clase* de una distribución de frecuencias se calcula haciendo la diferencia entre el límite superior y el límite inferior de la clase correspondiente.
- 48. Los *polígonos de frecuencias* sólo se pueden utilizar para representar las distribuciones de frecuencias relativas.
- 49. El polígono de frecuencias construido con las frecuencias simples *absolutas*, tiene la misma forma que el polígono de frecuencias simples *relativas* construido a partir del mismo conjunto de datos, pero con una escala diferente en los valores del eje vertical.
- 50. La representación gráfica de la distribución de frecuencias acumuladas mediante una poligonal, se conoce con el nombre de *ojiva*.

- 51. La *ojiva* es una representación gráfica que se puede construir utilizando frecuencias acumuladas absolutas o relativas, indistintamente.
- 52. La *ojiva* se construye uniendo los puntos dados por los pares ordenados (punto medio de clase; frecuencias acumuladas de clase), con trazos rectos que dan lugar a una poligonal.
- 53. En una distribución de frecuencias, la notación *F_{ri}* se refiere a las frecuencias de clase relativas acumuladas.
- 54. Dada la representación gráfica de una *ojiva*, a partir de la misma es posible reconstruir los datos originales exactos con los que se construyó la misma.
- 55. Si la lectura de la *ojiva* para la variable en estudio en la clase (12 ; 14] de una tabla de frecuencias arroja el valor 30%, debe interpretarse que el 30% de los datos están comprendidos en el intervalo (12 y 14].
- 56. Si la lectura de la *ojiva* para la variable en estudio en la *clase* (2,5 ; 3,5] de una tabla de frecuencias arroja el valor 60%, debe interpretarse que el 40% de los datos son mayores que 3,5.
- 57. Para describir en palabras *el patrón de comportamiento* de los datos, se puede hacer referencia a la simetría o asimetría de la distribución, a la presencia o no de *modas* en la distribución, así como al lugar en que tienden a agruparse los datos en la escala de la variable.

En la Tabla 1 se presenta la **distribución de frecuencias** para el peso, en gramos, de 35 monedas de diez centavos.

Tabla 1. Distribución de frecuencias para el peso de 35 monedas de diez centavos.

Clase	Límites de Clase		Punto Medio	Frecuencias Absoluta	Frecuencias Simples Relativa	Frecuencias Acumuladas	
	(Inferior	Superior]				Absoluta	Relativa
1	(2,13	2,16]	2,145	2	0,0571	2	0,0571
2	(2,16	2,19]	2,175	3	0,0857	5	0,1429
3	(2,19	2,22]	2,205	5	0,1429	10	0,2857
4	(2,22	2,25]	2,235	15	0,4286	25	0,7143
5	(2,25	2,28]	2,265	8	0,2286	33	0,9429
6	(2,28	2,31]	2,295	1	0,0286	34	0,9714
7	(2,31	2,34]	2,325	1	0,0286	35	1,0000

- 58. El 14,29% de las monedas de la muestra pesó más de 2,19 gramos, pero no superó los 2,22.
- 59. El 71,43% de las monedas de la muestra tiene un peso que no pasa de 2,25 gramos.
- 60. Hay 10 monedas en la muestra cuyo peso está por encima de los 2,25 gramos.
- 61. Ocho monedas de la muestra tienen un peso que en la distribución de frecuencias queda representado por el valor 2,265 gramos.
- 62. Las frecuencias de clase simples relativas de la Tabla 1 están expresadas en porcentaje.

- 63. De acuerdo a la información de la Tabla 1, la moneda más liviana de la muestra pesa 2,13 gramos.
- 64. El número de clases que se ha adoptado en la Tabla 1 concuerda con el propuesto por la fórmula de Sturges: $k = 1 + 3,3 \log n$.
- 65. El número de clases que se ha adoptado en la Tabla 1 concuerda con el propuesto por la fórmula \sqrt{n} .
- 66. Uno de los puntos de la representación gráfica de la *ojiva* correspondería al par ordenado $(x ; F)$: (2,25 ; 33).
- 67. En la muestra se observaron 8 monedas con un peso igual a 2,265 gramos.

2. Descripción de un conjunto de datos: Métodos numéricos.

- 68. La *media aritmética* de un conjunto de datos siempre coincide con alguno de los valores centrales del conjunto de valores observados.
- 69. La *media aritmética* siempre está comprendida entre los valores máximo y mínimo observados.
- 70. La *media aritmética* resulta siempre la mejor medida de tendencia central de un conjunto de datos numéricos.
- 71. En todo conjunto de datos numéricos, la *media* es un valor mayor o igual que cero.
- 72. La *media aritmética* es la mejor medida de tendencia central de un conjunto de datos categóricos.
- 73. Si la *media* o *promedio* de las calificaciones de un examen de Estadística, en la escala del cero al diez, resulta exactamente igual a diez puntos, el rango de tales calificaciones debe ser igual a cero.
- 74. Doce alumnos rinden un examen de Estadística, son calificados en la escala del cero al diez, y la *mediana* de las calificaciones es igual a seis. En tales condiciones, podría ocurrir que menos de cinco alumnos obtuvieran una calificación de seis puntos.
- 75. La suma de las desviaciones respecto de la media aritmética es siempre igual a cero.
- 76. Dado un conjunto numérico de datos de tamaño $n > 1$, la mediana puede o no existir.
- 77. La mediana es una medida de tendencia central sensible a los datos apartados de la muestra.
- 78. La mediana del siguiente conjunto de datos $\{2, 5, 7, 1, 3\}$ es igual a 7.
- 79. La moda puede no existir y cuando existe no necesariamente es única.
- 80. Dado un conjunto de mediciones resultantes de un experimento, la *moda* puede no coincidir con alguno de los valores observados.
- 81. La moda del siguiente conjunto de datos $\{3, 3, 3, 3, 3\}$ es igual a 3.
- 82. El valor de la media aritmética de un conjunto de datos categóricos es siempre menor que la moda de los mismos.

- 83. En todo conjunto de datos numéricos, la moda es un valor mayor o igual que cero.
- 84. Si un conjunto de datos no tiene moda, debe interpretarse que el valor numérico de la moda es igual a cero.
- 85. La *moda* es una medida de tendencia central que puede calcularse tanto para datos numéricos como para datos categóricos.
- 86. Si se tiene un conjunto de datos resultantes de medir la temperatura en el Parque General San Martín a la hora 8, podría suceder que se observe más de una *moda*.
- 87. La mediana es una medida de la variabilidad de un conjunto de datos numéricos.
- 88. La media aritmética es la medida que mejor describe la posición central del siguiente conjunto de datos $\{1; 2; 3; 3; 1; 2; 2; 1; 3; 12\}$.
- 89. En el siguiente conjunto de datos $\{-2, -1, 0, 1, x\}$, x podría asumir un valor tal que la *media* sea menor que la *mediana*.
- 90. Es suficiente calcular las medidas de tendencia central de una muestra, para proporcionar un resumen apropiado y acabado del conjunto de datos del cual proviene.
- 91. El *rango* del conjunto de datos siguiente $\{-6, -2, 0, 2, 6\}$ es igual a cero.
- 92. El *rango* de un conjunto de datos, siempre y sin restricción alguna, es un valor mayor o igual que cero.
- 93. Si el *rango* del conjunto de datos siguiente $\{1, 2, 3, x, 3, 2, 1\}$ es igual a tres, el valor de x sólo podría asumir el valor cero.
- 94. El *rango* es una medida *pobre* de la variabilidad, en particular si el tamaño de la muestra es grande; considera sólo los valores extremos y no nos dice nada acerca de la distribución de los valores intermedios.
- 95. La *varianza* del siguiente conjunto de datos $\{1, 1, 1, 1, 1\}$ es igual a 1^2 .
- 96. Cuando la *desviación estándar* de un conjunto de datos numérico es menor que cero, debe interpretarse que todos los datos son menores que la media aritmética.
- 97. La *desviación estándar* de un conjunto de datos, nunca puede resultar mayor que la *media* del mismo conjunto de datos.
- 98. Si la *desviación estándar* de la estatura de los alumnos de la Universidad es igual a 9 centímetros y la *desviación estándar* del promedio de calificaciones de los mismos alumnos es de 3 puntos, se debe concluir que la dispersión de las calificaciones es menor que la dispersión de las estaturas.
- 99. Si el rendimiento de un grupo de alumnos que es evaluado en Estadística resulta óptimo, digamos que todos obtienen por lo menos ocho puntos sobre diez, nada impide que la *desviación estándar* de las calificaciones resulte igual a 4 puntos.
- 100. Si la *unidad de medida* de la *desviación estándar* de una variable se expresa en metros, la *varianza* lo estará en metros cuadrados.
- 101. El *coeficiente de variación* permite comparar la dispersión o variabilidad de conjuntos de datos diferentes, incluso medidos en unidades diferentes.
- 102. El *coeficiente de variación* de cualquier conjunto de datos numéricos, expresado en porcentaje, está comprendido entre cero y cien.

- 103. Si la *desviación estándar* del caudal del *Río X* es de $45 \text{ m}^3/\text{s}$ y la *desviación estándar* del caudal del *Río Y* es de $240 \text{ m}^3/\text{s}$, se debe concluir que los caudales del *Río Y* están más dispersos que los del *Río X*.
- 104. Si una distribución tiene *sesgo positivo*, es asimétrica a derecha.
- 105. En cualquier conjunto de datos numéricos distribuidos simétricamente, media, mediana y moda son coincidentes.
- 106. Si una distribución de frecuencias de clase relativas resulta simétrica, la distribución de frecuencias acumuladas también lo será.
- 107. Si el tercer cuartil de un conjunto de datos observados es igual a 35, el 25% de los datos del conjunto es mayor que 35.
- 108. Si el valor del sexto decil de un conjunto de datos es igual a 8, significa que la sexta parte de los datos son iguales o inferiores a 8.
- 109. El percentil cincuenta de un conjunto de datos siempre coincide con el segundo cuartil.
- 110. Dado el conjunto de datos $\{1, 1, 1, 1, 1, 3, 5, 5, 5, 5, 5\}$ se cumple que el segundo decil es menor que el tercer cuartil.
- 111. Si se sabe que el percentil diez de un conjunto de datos es igual a 10, el primer decil será igual a 1.
- 112. En algunos conjuntos de datos, podría encontrarse que el percentil 22 resulte mayor que el cuartil inferior.
- 113. Si un fabricante de puertas para viviendas debe decidir qué altura darle a las mismas para una producción estándar, se le debe sugerir que adopte para las puertas, una altura igual a la estatura media de las personas adultas del mercado en el que se venderán dichas puertas.
- 114. Las estadísticas obtenidas de las muestras nos proporcionan información acerca de la tendencia central de los datos y de su dispersión, mientras que la presentación gráfica de los datos agrega información adicional en términos de imagen.
- 115. El gráfico de caja y extensión es una representación que muestra, para muestras razonablemente grandes, el centro de la localización, la variabilidad y el grado de asimetría de los datos.
- 116. Los gráficos de caja y extensión no permiten realizar comparaciones visuales entre muestras.
- 117. Los datos apartados (valores extremos) se deben identificar específicamente tanto en los gráficos de caja y extensión como los histogramas de frecuencias.
- 118. Tres de los datos necesarios para construir un gráfico de caja y extensión son: el primer cuartil, la mediana y el percentil setenta y cinco.
- 119. Datos apartados son aquellos que se encuentran por encima del tercer cuartil y por debajo del primer cuartil, más allá de 1,5 veces el rango intercuartil.
- 120. Los gráficos de caja y extensión NO proporcionan información sobre la variabilidad de los datos.

- _ 121. El gráfico de caja múltiple se puede utilizar para comparar la misma variable en muestras distintas.
- _ 122. Si el gráfico de caja es perfectamente simétrico, la varianza del conjunto de datos con que se construyó es igual a cero.
- _ 123. En el gráfico de caja, la caja siempre encierra exactamente el 50% de las observaciones.
- _ 124. Si se tiene un conjunto de cuarenta datos numéricos, el diagrama de tallos y hojas ofrece una información más detallada que el histograma de frecuencias.
- _ 125. Si se tiene un conjunto de cuarenta mil datos numéricos, el diagrama de tronco y hojas resultaría una representación más apropiada que el histograma de frecuencias, ya que da una información más detallada.
- _ 126. El *valor Z* debe estar comprendido entre -1 y $+1$.
- _ 127. Si el valor *Z* que le corresponde a una observación particular de la muestra, x , es negativo, debe interpretarse que el valor de x es menor que cero.
- _ 128. Si Pedro rindió una prueba de Estadística y obtuvo una calificación tal que el *valor Z* correspondiente es igual a 2, debe interpretarse que Pedro aprobó el examen.
- _ 129. Pedro y Juan son estudiantes de la clase de Estadística. Pedro tiene una estatura que coincide con la estatura promedio del grupo, mientras que a la estatura de Juan le corresponde un *valor Z* igual a $-2,95$. Debe interpretarse entonces que Juan tiene una estatura apenas por debajo de la de Pedro.
- _ 130. La *media* de los *valores Z* de un conjunto de datos numéricos es siempre igual a 1.
- _ 131. La *desviación estándar* de los *valores Z* de un conjunto de datos numéricos, puede arrojar un valor comprendido entre 0 y 1.

3. Aspectos éticos

- _ 132. Debe distinguirse entre una mala presentación de los datos y una presentación que carece de ética.
- _ 133. La conducta NO ÉTICA se da cuando el analista oculta hechos a propósito y/o distorsiona tablas o gráficos; también, cuando no incluye los hallazgos pertinentes.
- _ 134. No entregar un trabajo en término por estar enfermo, es una conducta NO ÉTICA.