

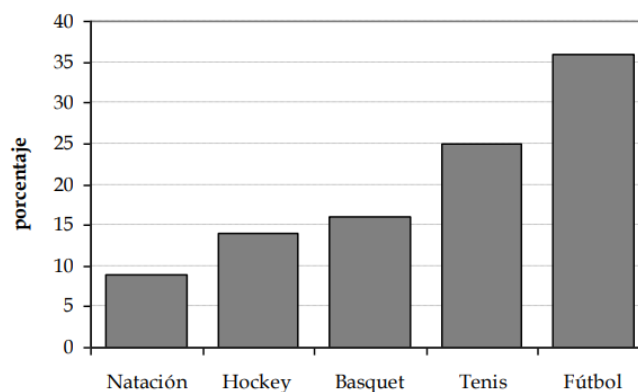
1. Investigadores de la Escuela de Salud Pública de la facultad de Ciencias Médicas (UNR) realizaron un estudio en el que encontraron que tomar gaseosas de forma regular podría asociarse a una mayor frecuencia de fracturas de huesos (vinculada a la osteoporosis, un problema frecuente en las mujeres). Para llegar a esta conclusión, encuestaron en la Universidad a 2.622 mujeres físicamente activas y las clasificaron según si tenían (o no) el hábito de tomar gaseosas de forma periódica. El resultado: aquellas que regularmente bebían gaseosas presentaron el doble de probabilidad de sufrir fracturas óseas que aquellas que tomaban gaseosas de forma esporádica.
 - a) ¿Se trata de un estudio observacional o experimental?
 - b) ¿Cuál fue la variable explicativa?
 - c) ¿Cuál fue la variable respuesta?
 - d) En base a los resultados, ¿sería correcto afirmar que tomar gaseosas produce fracturas de huesos?
 - e) Suponga que las 2.622 mujeres incluidas en el estudio fueron seleccionadas del siguiente modo: “De una lista de todos los eventos deportivos femeninos de alto rendimiento disponibles en la Universidad (tenis, básquet, etc.) se seleccionaron aleatoriamente 3 eventos. Todas las mujeres que participaron en esos 3 eventos fueron contactadas para responder la encuesta.” ¿Qué técnica de muestreo se aplicó?
 - f) La muestra del enunciado, ¿es representativa de la población de mujeres de Argentina? ¿Y la del ítem e)?
2. En un programa de radio que se escucha a nivel nacional se invitó a las/los oyentes a contestar la siguiente pregunta: “Si pudiera volver el tiempo atrás, ¿volvería a tener hijos?”. De más de 10.000 respuestas, el 70% dijo que no. ¿Qué muestra esto? Evalúe cada una de las siguientes afirmaciones. Justificar.
 - a) Es una muestra legítima elegida al azar entre todos los que escuchan ese programa.
 - b) El 70% de la población argentina desearía no haber tenido hijos.
 - c) El 70% de los oyentes de esa radio se arrepiente de haber tenido hijos.
 - d) Para sacar una conclusión, es necesario separar las respuestas entre hombres y mujeres.
 - e) No se pueden extraer conclusiones porque la encuesta puede presentar sesgo por respuesta voluntaria.
 - f) Hubiese tenido más sentido tomar una muestra aleatoria de las 10.000 respuestas para sacar conclusiones.
 - g) Si bien se trata de una muestra sesgada, el tamaño de la muestra es lo suficientemente grande como para corregir el sesgo en la estimación.
3. Se quiere realizar una encuesta entre los alumnos de una escuela secundaria, de 2.500 alumnos (500 alumnos por cada año, de 1ro a 5to), utilizando una muestra de tamaño 100. El propósito de la encuesta es determinar si a las/os alumnas/os les interesa discutir el siguiente tema: “Debe reducirse la edad de imputabilidad penal para los menores de edad, que establece la ley nacional 22.278, a dieciséis años de edad, como respuesta al incremento en la cantidad de delitos graves cometidos por jóvenes y adolescentes”. Indicar cuál es el tipo de muestreo realizado en cada caso:
 - a. Cada alumno escribe su nombre en un papel, lo pone en una bolsa y el director elige 100 papeles;

- b. A cada alumno se le asigna un número entre 1 y 2.500 y se seleccionan generando 100 números al azar de cuatro dígitos utilizando algún programa de computación.
 - c. A cada alumno se le asigna un número entre 1 y 2.500, y se elige 1 de cada 25 alumnos.
 - d. Se eligen al azar algunas divisiones y se incluye en la muestra a los 20 alumnos de cada división.
 - e. Se eligen al azar 60 alumnos de los primeros 3 años y 40 alumnos de los últimos dos años.
- 4. Como parte del estudio de la biosfera del Bosque de los Constituyentes de Rosario, un equipo de investigadores necesita juntar información sobre la presencia de nidos de aves. El bosque fue dividido en $N=80$ sectores de aproximadamente igual tamaño. Debido al tiempo y costo, los investigadores decidieron seleccionar 5 de estos sectores al azar y enviar a un equipo de investigadores a cada sector para su observación. Necesitamos seleccionar una muestra simple al azar de tamaño $n=5$ de una población de $N=80$.
 - a. ¿Cómo sería el marco muestral? Bosqueje.
 - b. ¿Cuál es la probabilidad que tiene cada sector de ser incluido en la muestra?
 - c. ¿Cuál es la probabilidad que tiene cada muestra de tamaño $n=5$ de ser seleccionada?
 - d. ¿Qué mecanismos generadores de aleatoriedad podríamos utilizar para seleccionar una muestra simple al azar?
 - e. Seleccionar una muestra utilizando la tabla de números aleatorios (fila 10).
 - f. Seleccionar una muestra utilizando otro mecanismo.
- 5. En cada uno de los siguientes enunciados indicar: i) si se trata de un estudio observacional o experimental, ii) las variables, indicando -si es posible- respuestas, explicativas y de confusión, iii) la unidad de análisis y iv) la cantidad de datos recolectados (n).
 - a. Se pretende estudiar si el azufre produce una disminución del número de cierto tipo de parásitos de la papa. Para ello se divide el campo en 24 parcelas de igual tamaño. Interesa comparar 3 cantidades distintas de azufre en 2 estaciones del año (primavera y otoño). En cada estación del año y con cada cantidad de azufre se repite el experimento en 4 parcelas que son asignadas al azar entre las 24 parcelas preparadas para el experimento. En cada parcela se mide el número de parásitos de la papa.
 - b. El Ministerio de Salud de la Nación junto al INDEC lleva a cabo la Encuesta Nacional de Factores de Riesgo (ENFR), la cual proporciona información válida, confiable y oportuna sobre factores de riesgo, procesos de atención en el sistema de salud y principales enfermedades no transmisibles en la población argentina. El diseño muestral es multietápico, logrando encuestar a personas de 18 años y más de Argentina para obtener la información necesaria. Las variables temáticas consideradas en la encuesta son la actividad física, la alimentación, el peso corporal y el tabaco, así como el uso de casco o cinturón de seguridad como indicador de seguridad vial. Se releva además la presión arterial, colesterol, glucemia y diabetes.
 - c. Es sabido que entre los factores que influyen en la determinación del salario de ejecutivos de empresas radicadas en Brasil figuran el nivel educativo alcanzado y los años de experiencia. Es por ello, que se desea realizar una investigación para verificar si una situación similar se observa en Argentina. Se selecciona

una muestra aleatoria de 100 ejecutivos a los cuales se les consulta sobre su salario, su nivel educativo y los años de experiencia.

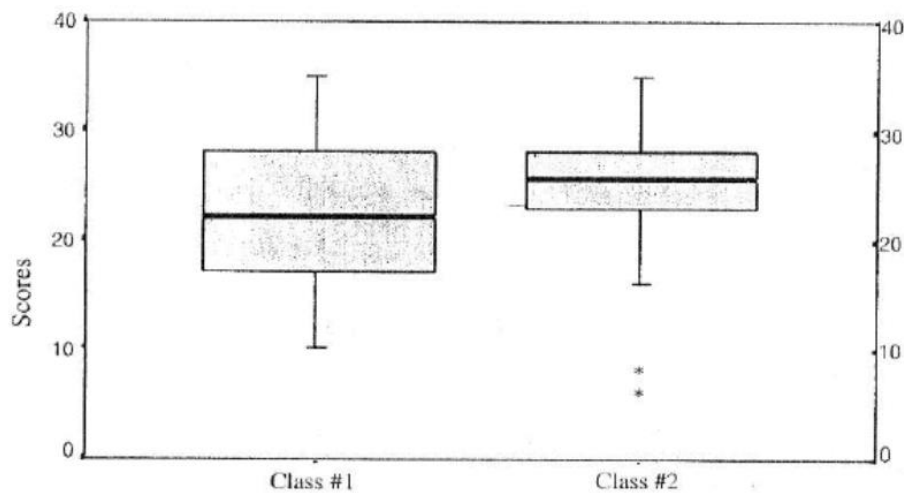
- d. Un investigador desea estudiar los efectos de los medicamentos A y B sobre el recuento de linfocitos en ratones (expresados en miles por ml³ de sangre) comparando A, B y el placebo P. Al diseñar el experimento se piensa que los ratones de una misma camada darán respuestas más homogéneas que ratones de camadas diferentes. Por tal motivo se utilizan en el experimento 7 camadas de 3 ratones cada una.

6. El canal de deportes TyC Sports realizó una investigación para conocer cuál es el deporte que los argentinos prefieren ver por televisión. Los resultados de una encuesta realizada a 36 televidentes han sido resumidos en el gráfico de barras que se presenta a continuación.



- a. ¿Qué tipo de variable es “deporte favorito”? Elija la respuesta correcta.
- Cualitativa
 - Cuantitativa y discreta
 - Cuantitativa y continua
- b. ¿Cuál es el modo de la distribución de la variable “deporte favorito”?
- c. ¿Aproximadamente, qué porcentaje de televidentes prefiere ver básquet?
- d. ¿Cuántos televidentes respondieron tenis como el deporte preferido para ver por televisión?
- e.Cuál(es) de la(s) siguiente(s) descripción(es) es(son) apropiada(s) para la distribución de la variable “deporte preferido”.
- Simétrica
 - Bimodal
 - Asimétrica a la derecha
 - Uniforme
 - Asimétrica a la izquierda
 - Ninguna de estas opciones es correcta

7. A continuación, se presentan los diagramas de caja de las notas obtenidas por los alumnos de dos comisiones de la Facultad de Ciencias Económicas y Estadística en el primer parcial de la materia Estadística.



- a. El alumno que obtuvo la nota más baja en el primer parcial de Estadística pertenece a la comisión 1.

Verdadero

Falso

No se puede responder

- b. La comisión 2 tiene un número de alumnos inferior a la comisión 1.

Verdadero

Falso

No se puede responder

- c. La distribución de las notas del primer parcial de Estadística para los alumnos de la comisión 1 es simétrica.

Verdadero

Falso

No se puede responder

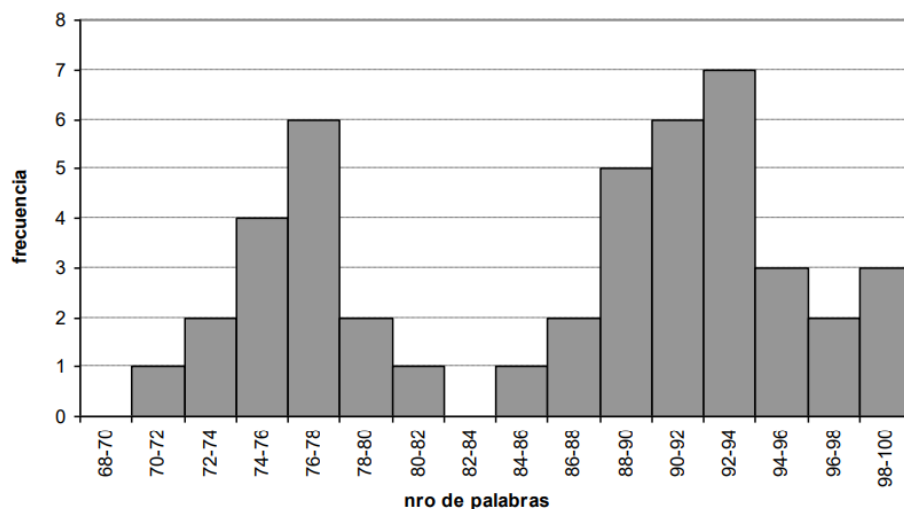
- d. Más del 75% de los estudiantes de la comisión 2 obtuvieron notas en el primer parcial de Estadística mayores que la nota mediana de los alumnos de la comisión 1.

Verdadero

Falso

No se puede responder

8. Un psicólogo ha desarrollado una nueva técnica para mejorar la memoria. Para probar el método contra otros métodos estándares, seleccionó a 45 estudiantes de la escuela secundaria al azar y a cada uno le enseñó la nueva técnica. Los datos sobre el número de palabras memorizadas correctamente fueron resumidos en el siguiente gráfico.



Nota: En el histograma, la primera clase representa el intervalo [68, 70). Se sabe que

$$\sum_{i=1}^{45} x_i = 3873.$$

- a. ¿Qué proporción de estudiantes memorizaron correctamente al menos 94 palabras?
- b. Describa la forma de la distribución del número de palabras memorizadas por estos estudiantes.
- c. ¿Es apropiado informar el promedio como una medida de tendencia central de estos datos? Explique.
- d. El histograma muestra evidencia de la presencia de dos subgrupos que no deberían ser agregados (o combinados). ¿Se vería esto mismo a través de un diagrama de caja o box plot para las 45 observaciones? Explique.
- e. ¿Puede usted calcular el valor exacto del rango para estas 45 respuestas? Si su respuesta es afirmativa, calcúlelo. En otro caso, explique por qué no.

9. Las edades de los sujetos sometidos a un estudio de reproducibilidad son las siguientes:

22 22 23 24 25 25 26 27 28 29 29 29 29 29 31 31 32 33 34 35 35
35 36 38 39 39 42 42 44 44 45 45 45 47 48 52 59 66 67 69 69

- a. Defina y clasifique la variable en estudio.
 - b. Guarde los valores correspondientes a la variable edad en un vector.
 - c. Ensaye comandos R a fin de obtener:
 - c1) Tamaño de la muestra.
 - c2) Media, mediana, moda y cuartiles.
 - c3) Rango, rango intercuartílico. Desviación estándar. Variancia.
 - c4) Percentil 10 y percentil 50.
 - c5) Mínimo y máximo y cuartiles, todos mediante un único comando.
 - c6) Diagrama de caja.
 - c7) Diagrama tallo-hoja.
 - c8) Histograma.
 - d. Interprete los resultados obtenidos en el ítem c).
10. Se realiza un estudio acerca de los efectos del tabaquismo sobre los patrones de sueño. La medición que se observa es el tiempo, en minutos, que toma quedar dormido. Se obtienen estos datos:

Fumadores:

69.3 56.0 22.1 47.6
53.2 48.1 52.7 34.4
60.2 43.8 23.2 13.8

No fumadores:

28.6 25.1 26.4 34.9
29.8 28.4 38.5 30.2
30.6 31.8 41.6 21.1
36.0 37.9 13.9

- a. Encuentre la media de la muestra para cada grupo.
- b. Encuentre la desviación estándar de la muestra para cada grupo.
- c. Usando un diagrama de caja comparativo grafique los conjuntos de datos fumadores vs. no fumadores.
- d. Comente qué clase de impacto parece tener el hecho de fumar sobre el tiempo que se requiere para quedarse dormido.

11. Analice en forma completa el conjunto de datos 'chickwts' (paquete:datasets)
12. Analice el conjunto de datos del archivo arbol.xlsx. (adjunto)