

1. 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).
2. 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:

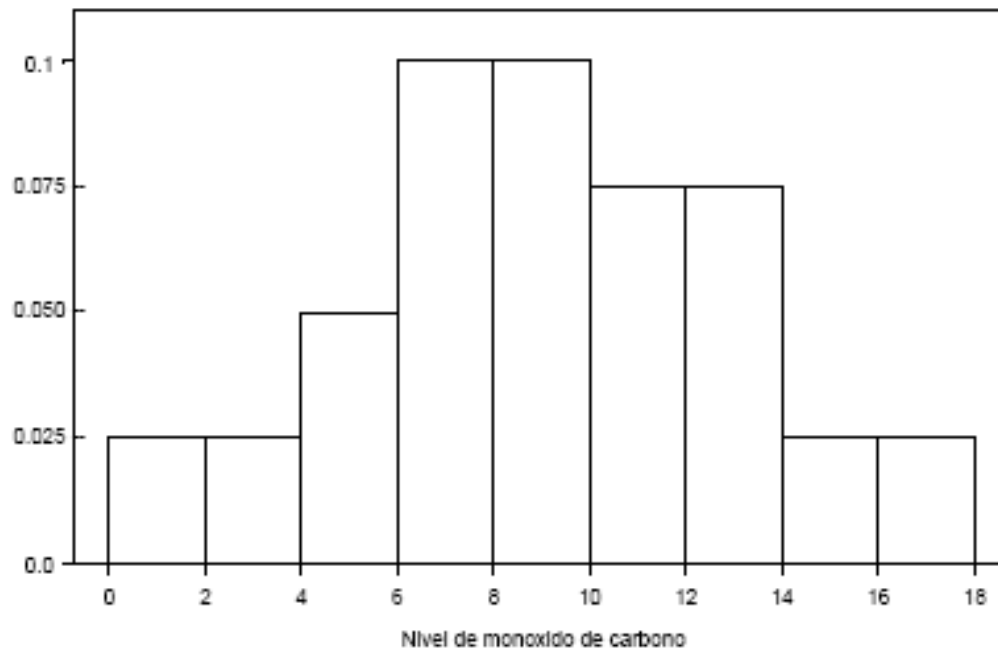
Fumadores:

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.
3. En un debate sobre contaminación ambiental se está evaluando la concentración de monóxido de carbono en las esquinas de la ciudad en los momentos de mayor tráfico (los viernes a la tarde). Interesa estimar p , la probabilidad de que una esquina de la ciudad elegida al azar presente valores de monóxido inferiores a 12 ppm (partes por millón). Para ello se recolecta una muestra de 120 valores correspondientes al nivel de monóxido de carbono en esquinas de la ciudad elegidas al azar un viernes a la tarde. Los datos obtenidos se representan en el siguiente histograma, en el cual lamentablemente se omitió el rótulo del eje de las ordenadas.
- a. Especifique el rótulo omitido
 - b. Construya a partir del histograma la tabla de distribución de frecuencias, interpreta la cuarta fila y calcule todas las medidas descriptivas que conoce y describa la forma de la distribución.



4. Analice en forma completa el conjunto de datos 'chickwts' (paquete:datasets)
5. Analice el conjunto de datos del archivo arbol.xlsx. (adjunto)
6. Investigadores de la Escuela de Salud Pública de Harvard 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). 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) Suponga que las 2.622 mujeres incluidas en el estudio fueron seleccionadas del siguiente modo: “De una lista de todos los eventos deportivos femeninos 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?
7. En un programa de radio 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) No se pueden extraer conclusiones porque la encuesta puede presentar sesgo por respuesta voluntaria.
 - b) No se puede decir nada sin saber las características de los oyentes.
 - c) Para sacar una conclusión, es necesario separar las respuestas entre hombres y mujeres.
 - d) Hubiese tenido más sentido tomar una muestra aleatoria de las 10.000 respuestas para sacar conclusiones.
 - e) Es una muestra legítima elegida al azar entre todos los que escuchan ese programa.
 - f) 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.
8. 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.
9. En cada uno de los siguientes enunciados indicar: i) si se trata de un estudio observacional o experimental, ii) la variable respuesta -en caso de que haya alguna-, iii) la unidad de análisis, iv) el factor o factores en estudio (variable/s explicativa/s) con sus respectivos niveles -si es que se informan- y v) 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 ml^3 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.