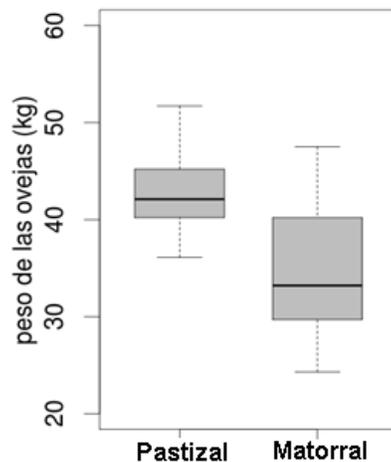


ESTADÍSTICA GENERAL

Ejercitación de repaso

Ejercicio 1. En una estancia de Chubut se registró el peso de las ovejas de dos muestras aleatorias de 200 ovejas cada una, la primera tomada en un potrero con vegetación de pastizal y la segunda en un potrero con vegetación de matorral de arbustos. A continuación se presentan gráficos de caja y bigotes que representan las distribuciones de frecuencias de los valores obtenidos:



1. ¿Cuáles son las poblaciones, las unidades muestrales y las muestras en este caso?
2. Aproximadamente, ¿cuánto valen las medianas de los pesos de las ovejas en las muestras obtenidas en cada potrero? y ¿cuánto vale la frecuencia relativa de ovejas con peso mayor que 40 kg en cada una de las dos muestras?
3. ¿En cuál tipo de vegetación se observó mayor dispersión en los pesos de las ovejas? Justificar la respuesta con una medida apropiada extraída del gráfico.

Ejercicio 2. Un ingeniero forestal observó 2000 árboles de la especie *Nothofagus pumillo* (lenga) de un bosque de la provincia de Río Negro y los clasificó según su diámetro a la altura del pecho en grandes (d.a.p. > que 10 cm) y chicos (d.a.p. ≤ 10 cm). Además registró si cada árbol estaba colonizado por la planta hemiparásita *Misodendrum punctulatum*. Sus registros indican que 1400 de los árboles correspondían a la categoría de tamaño chico y los restantes a la categoría de tamaño grande; el 36 % de los árboles con d.a.p. menor o igual que 10cm y 300 de los mayores estaban colonizados por el hemiparásito. Definamos ahora el experimento aleatorio que consiste en tomar al azar uno de los 2000 árboles referidos y determinar a qué clase de tamaño y a qué clase de infección corresponde:

1. ¿Qué eventos simples componen el espacio muestral de este experimento? O dicho de otro modo ¿Cuáles son los posibles resultados del experimento?
2. Señalar dos eventos mutuamente excluyentes en dicho espacio. ¿Cuál es la probabilidad de cada uno? ¿Cuál es la probabilidad de que ocurra uno o el otro?
3. Señalar dos eventos no mutuamente excluyentes en dicho espacio. ¿Cuál es la probabilidad de cada uno? ¿Cuál es la probabilidad de que ocurra uno o el otro?
4. ¿Cuál es la probabilidad de que un árbol a tomar al azar corresponda a la categoría de tamaño d.a.p. > 10cm?
5. ¿Cuál es la probabilidad de que un árbol de la categoría de tamaño d.a.p. > 10cm a tomar al azar no esté colonizado por *Misodendrum punctulatum*?

6. ¿El tamaño y la infección con *Misodendrum punctulatum* de un árbol a tomar al azar son estadísticamente independientes?

Ejercicio 3. La Mancha Ojo de Rana, causada por el hongo *Cercospora sojina*, es una enfermedad de la soja que produce lesiones que disminuyen su función fotosintética de las hojas de esta especie. La intensidad del ataque de esta enfermedad se puede evaluar mediante el número de manchas por hoja. Si en un gran cultivo de soja la distribución de frecuencias relativas de hojas con diferente número de manchas por hoja es la que se indica en la siguiente tabla:

<i>manchas por hoja</i>	0	1	2	3	4	5	> 5
<i>Frecuencia Relativa</i>	0,05	0,18	0,28	0,24	0,15	0,10	0

Definamos X como el número de manchas en una hoja a tomar al azar de este cultivo.

1. ¿Por qué X es una variable aleatoria? ¿Cuál es el experimento aleatorio que la genera? ¿Qué tipo de variable aleatoria es? ¿Qué valores puede tomar?
2. ¿Cuál es el valor de la probabilidad de que una hoja a tomar al azar de este cultivo tenga más de 1 pero no más que 4 manchas?
3. ¿Cuánto vale la esperanza del número de manchas en una hoja a tomar al azar de este cultivo? ¿A qué característica de la población de referencia corresponde este valor?
4. ¿Cuánto vale la varianza del número de manchas en una hoja a tomar al azar de este cultivo? ¿A qué característica de la población de referencia corresponde este valor?

Ejercicio 4. Sea U una variable aleatoria que toma valores en el intervalo $[0, 2]$ y cuya función de densidad de probabilidad es:

$$f(u) = 0,5 u, \text{ para } 0 \leq u \leq 2; \text{ y } f(u) = 0, \text{ para cualquier otro valor de } u$$

1. Graficar la función de densidad de probabilidad.
2. Calcular los valores de los percentiles 25, 50 y 75 de la distribución de probabilidad de la variable U y señalarlos en el gráfico (usar para ello la fórmula del área de un triángulo).
3. Calcular el valor de $E(U)$, la esperanza de la variable aleatoria U .
4. Calcular la probabilidad de que la variable aleatoria U tome un valor menor que $E(U)$.

Ejercicio 5. Si el tamaño de una manzana a tomar al azar de la línea de empaque del Alto Valle de Río Negro es una variable aleatoria X con distribución Normal con $\mu = 78$ mm y $\sigma = 4$ mm.

1. ¿Qué experimento aleatorio genera la variable X ? ¿Qué población de referencia involucra?
2. ¿Qué tipo de variable es? ¿Qué valores puede tomar?
3. ¿Cómo se define la esperanza de X ? ¿Qué valor numérico tiene? ¿Con qué característica de la población coincide dicho valor?
4. ¿Cómo se define la varianza de X ? ¿Qué valor numérico tiene? ¿Con qué característica de la población coincide dicho valor?
5. ¿Cuál es la probabilidad de que una manzana a tomar al azar tenga un tamaño entre 76 y 80 mm?
6. ¿Cuál es el tamaño que es superado por el 90% de las manzanas?
7. Si se decide tomar una muestra aleatoria de 25 manzanas, ¿qué distribución de probabilidad tiene la media muestral de los tamaños?
8. ¿Cuál es la probabilidad de que dicha media muestral tome un valor entre 76 y 80 mm?

Ejercicio 6. Un productor familiar de la zona de Cañuelas que tiene 16 vacas lecheras desea incrementar su producción de leche aumentando el número de vacas de su tambo. Para ello decide conservar las hembras que nazcan un año para incorporarlas en el rodeo. Si la

probabilidad de que cada vaca quede preñada en el año es de 0,96 y dado que es igualmente probable que una vaca preñada produzca una hembra o un macho.

1. ¿Cuánto vale la probabilidad de que una vaca quede preñada de una hembra?
2. ¿Qué distribución de probabilidad tiene el número de hembras por nacer en el año?
3. ¿Cuánto vale la probabilidad de que ese año nazcan al menos 10 hembras?
4. ¿Qué valores tienen la esperanza y la varianza del número de hembras por nacer?

Ejercicio 7. En una investigación sobre el rendimiento del cultivo de arroz en la región del litoral fueron registrados los rendimientos en establecimientos que aplican 2 diferentes sistemas de riego A y B. Los datos generados están resumidos en la tabla:

Rendimiento (kg/ha)	Frecuencia Absoluta	
	Sistema A	Sistema B
(1500 , 2000]	10	4
(2000 , 2500]	50	10
(2500 , 3000]	30	40
(3000 , 3500]	10	40
(3500 , 4000]	0	6

1. Elaborar un gráfico adecuado para visualizar esta información
2. Calcular valores aproximados de la media y el desvío estándar de los rendimientos registrados para cada método de riego.
3. ¿Cuál de los métodos de riego resultó en promedio en mayores rendimientos?
4. ¿Cuál de los métodos de riego resultó en rendimientos más variables?
5. ¿Qué se puede decir respecto de la conveniencia de adoptar el método de riego B?

Ejercicio 8. Después de un terremoto el 30% de las víctimas de una zona rural contrajo disentería. En dicha zona, una enfermera de la Cruz Roja atiende a 6 víctimas del terremoto por turno. Si se puede dar por cierto que cada grupo de 6 víctimas constituye una muestra aleatoria de la población de víctimas:

1. ¿Cuál es la probabilidad de que en un turno, al menos una de las víctimas tenga disentería?
2. ¿Cuál es el promedio de personas con disentería por turno de atención?
3. ¿Cuál es el valor del desvío estándar del número de personas con disentería por turno de atención?

Ejercicio 9. En el rótulo de una partida de higos de la variedad King que entran a una planta elaboradora dice que tienen un peso medio = 116 g y un desvío estándar = 17 g. En la planta elaboradora, los higos más pesados que el percentil 80 se destinan a elaborar higos en almíbar y los restantes se destinan a producir mermelada. Suponiendo que lo que dice en el rótulo es cierto y además el peso de los higos es una variable aleatoria con distribución Normal,

1. ¿Cuál es la probabilidad de que el peso de un higo a tomar al azar de esta partida fuera destinado a producir mermelada?
2. ¿Cuál es el máximo peso de los higos destinados a producir mermelada?
3. ¿Cuál es la probabilidad de que entre 16 higos a tomar al azar de esta partida no menos de 2 tengan pesos apropiados para la producción de higos en almíbar?

Ejercicio 10. Imaginemos el experimento aleatorio que consiste en arrojar un dardo a un blanco de 1 m de radio y registrar la distancia entre el punto donde el dardo impacta y el centro del blanco (m). La distancia a registrar es una variable aleatoria que denominaremos X . Suponiendo que la función de densidad de probabilidad de X es la siguiente:

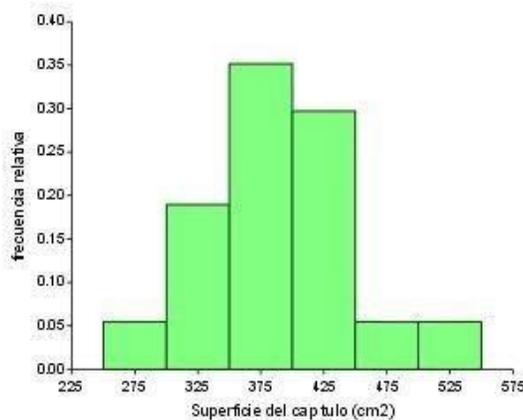
$$f(x) = \begin{cases} 2 \cdot (1 - x) & \text{para } x \in [0, 1] \\ 0 & \text{para cualquier otro valor de } x \end{cases}$$

1. Graficar la función de densidad de probabilidad de X .
2. ¿Cuánto vale la probabilidad de que el dardo impacte exactamente en el centro del blanco?
3. ¿Cuánto vale la probabilidad de que el dardo impacte a menos de 0,25 m del centro?
4. ¿Cómo se define la esperanza de la variable aleatoria X ? ¿Cuánto vale? ¿Cómo se interpreta?
5. ¿Cómo se define la varianza de la variable aleatoria X ? ¿Cuánto vale? ¿Cómo se interpreta?
6. Graficar la función de distribución de probabilidad acumulada de X y localizar el valor aproximado de la mediana.
7. ¿Cuál es el valor exacto de la mediana de la distribución de probabilidad de X ?

Ejercicio 11. En un lago conviven dos especies de pejerrey (A y B) en igual proporción. El 22% de los pejerreyes de la especie A y el 35% de los de la especie B están infectados por un protozoo intestinal.

1. ¿Cuál es la probabilidad de que un pejerrey a tomar al azar esté infectado?
2. ¿Cuál es la probabilidad de que un pejerrey a tomar al azar esté infectado y además pertenezca a la especie A?
3. ¿Cuál es la probabilidad de que un pejerrey infectado a tomar al azar pertenezca a la especie A?
4. Usando probabilidades condicionales, explicar por qué la infección con el protozoo intestinal no es estadísticamente independiente de la especie de pejerrey.

Ejercicio 12. El siguiente gráfico representa la distribución de frecuencias de la variable superficie del capítulo de 37 plantas de girasol del híbrido AH32 cultivadas en un lote con riego. La cantidad de agua de riego aplicada durante el ciclo del cultivo fue de 600mm.



1. ¿De qué tipo de gráfico se trata?
2. Con la información de este gráfico, construir un gráfico de frecuencias relativas acumuladas (aproximadas).
3. ¿Cuáles son los valores aproximados del primer cuartil, de la mediana y del tercer cuartil?
4. Otro lote fue regado con 800mm de agua. Si la superficie del capítulo fuera una característica plástica ¿qué diferencia sería esperable encontrar entre la distribución de frecuencias de la superficie de los capítulos de dicho lote y el que se presenta aquí?

Ejercicio 13. Un fitotecnista ensaya un nuevo híbrido de girasol para decidir si conviene introducirlo en el norte de Santa Fe. Para ello, se dispone a seleccionar al azar 25 lotes dentro de la región, cultivar en cada lote el híbrido en cuestión y registrar los rendimientos que obtenga en cada uno. Luego calculará la media aritmética de los rendimientos que registre (media muestral).

1. Identificar la población, la muestra y las variables aleatorias a registrar.
2. Explicar qué significa que las variables aleatorias a registrar son independientes e idénticamente distribuidas.
3. ¿Qué valor tendrá la esperanza de dichas variables aleatorias?
4. ¿Qué valor tendrá la varianza de dichas variables aleatorias?
5. Explicar por qué la media aritmética de los rendimientos a obtener es una variable aleatoria.
6. Explicar qué es la esperanza de dicha media muestral. ¿Cuál será su valor?
7. Explicar qué es la varianza de dicha media muestral. ¿Cuál será su valor?
8. ¿A qué modelo se aproxima la distribución de probabilidad de la media muestral a calcular?

Ejercicio 14. Si el rendimiento de un lote de girasol a tomar al azar en el norte de Santa Fe en tn/ha es una variable aleatoria con distribución aproximadamente Normal con $\mu = 2,5$ y $\sigma = 0,6$.

1. Identificar el experimento aleatorio que produce la variable aleatoria en cuestión.
2. ¿Cuánto vale la esperanza del rendimiento de un lote de girasol a tomar al azar en el sur de Santa Fe? ¿Por qué nos interesa este valor?
3. ¿Cuánto vale la varianza del rendimiento de un lote de girasol a tomar al azar en el sur de Santa Fe?
4. Nombrar dos posibles causas materiales de la varianza de los rendimientos del cultivo de girasol.
5. ¿Cuál es el valor de rendimiento superado por el 10% de los lotes de girasol del sur de Santa Fe?
6. ¿Cuál es el valor del percentil 75 de la distribución de frecuencias de los rendimientos de girasol en el sur de Santa Fe?
7. ¿Cuál es la probabilidad de que un lote a tomar al azar tenga un rendimiento superior a los 2500 kg/ha?
8. ¿Cuál es la probabilidad de la media aritmética de los rendimientos de 10 lotes de girasol a tomar al azar en el sur de Santa Fe supere los 2500 kg/ha?
9. ¿Cuál es la probabilidad de que un lote a tomar al azar tenga un rendimiento superior a los 2800 kg/ha? ¿Qué relación tiene esa probabilidad con la frecuencia relativa de lotes con rendimiento superior a 2800 kg/ha en la población formada por todos los lotes de girasol del sur de Santa Fe?
10. ¿Cuál es la probabilidad de que la media aritmética de los rendimientos de 10 lotes de girasol a tomar al azar en el sur de Santa Fe supere los 2800 kg/ha?

Ejercicio 15. En la Depresión de Salado (Prov. de Buenos Aires), el agua de los pozos que se usa para el ganado contiene concentraciones variables de sales. Un técnico planea estimar el promedio de los contenidos de sales de todos los pozos de dicha región a partir de los datos de una muestra aleatoria de 25 pozos de la región.

1. ¿En qué consiste el procedimiento que el técnico debe aplicar para obtener la muestra aleatoria? ¿Qué variables aleatorias registrará?
2. ¿Por qué la media muestral de los contenidos de sales del agua de los 25 pozos que se incluyan en la muestra es una variable aleatoria?
3. ¿Qué es la esperanza de la media muestral en este caso particular? ¿Qué valor tiene?
4. ¿De qué depende la probabilidad de que la media muestral a obtener tome un valor cercano al de la media poblacional que se planea estimar?

5. Si en verdad los contenidos de sales del agua de todos los pozos de la Depresión del Salado tienen promedio $\mu = 75$ mg /l y desvío estándar $\sigma = 22,8$ mg /l,
6. ¿Cuál es el valor aproximado de la probabilidad de que la media muestral que obtendrá el técnico tome un valor entre 64,8 y 85,2 mg/l?
7. ¿Qué tamaño de muestra debe tomar el técnico para que la probabilidad de que la media muestral difiera de la media poblacional en menos que 6 mg /l sea a lo sumo de 0,90?