

ESTADÍSTICA GENERAL

Ejercitación de repaso general

Ejercicio 1. Sea X una variable de interés cuya distribución de frecuencias en la población de referencia se ajusta a una distribución Normal con $\mu = 20$ y con $\sigma = 2$. Alguien que no conoce estos valores planea estimarlos a partir de una muestra aleatoria de 16 unidades muestrales de la población de referencia en cada una de las cuales registrará los correspondientes valores de X . Con los datos que obtenga, esta persona calculará los siguientes estadísticos:

$$\bar{X} = \sum_{i=1}^{16} \frac{X_i}{16}$$

$$S^2 = \sum_{i=1}^{16} \frac{(X_i - \bar{X})^2}{15}$$

1. Explicar qué es la esperanza de X y qué es la varianza de X .
2. ¿Cómo llaman los estadísticos \bar{X} y S^2 ?
3. Explicar por qué los estadísticos \bar{X} y S^2 son variables aleatorias.
4. Explicar qué es la esperanza de \bar{X} .
5. ¿Cuál es el valor de la esperanza de \bar{X} ?
6. Explicar qué es la varianza de \bar{X} .
7. ¿Cuál es el valor de la varianza de \bar{X} ?
8. ¿Cuál es la distribución de probabilidad de \bar{X} ?
9. ¿Cuál es la probabilidad de que \bar{X} tome un valor que no difiera de la esperanza de X en más que 1?
10. Si se obtienen 3 muestras aleatorias de 16 unidades en las que se registra la variable aleatoria X , ¿cuál es la probabilidad de que el estadístico \bar{X} correspondiente a exactamente una de las muestras tome un valor que no difiera de la esperanza de X en más que 1?
11. Explicar qué es la esperanza de S^2 .
12. ¿Cuál es el valor de la esperanza de S^2 ?
13. ¿Qué distribución de probabilidad tiene el siguiente estadístico?
$$\frac{\bar{X} - 20}{\sqrt{\frac{S^2}{16}}}$$
14. ¿Cuánto vale la siguiente probabilidad?
$$P \left[\frac{\bar{X} - 20}{\sqrt{\frac{S^2}{16}}} < 2,602 \right]$$
15. ¿Cuál es la forma genérica de un intervalo de confianza para μ ?
16. ¿Cómo se modifica la precisión del intervalo de confianza para μ si aumenta la varianza de X , si se cambia el nivel de confianza elegido, o si se aumenta el tamaño de la muestra?
17. ¿Cuál es el estadístico de prueba apropiado para poner a prueba la hipótesis nula $H_0: \mu \geq 25$?
18. ¿Cómo se calcula el valor p para poner a prueba la hipótesis nula $H_0: \mu \geq 25$?
19. ¿Cuál es, en este caso, la verdadera distribución de probabilidad del estadístico de prueba apropiado para poner a prueba la hipótesis nula $H_0: \mu \geq 25$? (Ayuda: ver página 153)

Ejercicio 2. 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 un cierto número de lotes agrícolas dentro de la región, cultivar en cada lote el híbrido en cuestión y registrar los rendimientos que obtenga en cada uno. Con los rendimientos que registre calculará la media muestral y la *varianza muestral* (estimador insesgado de la varianza poblacional) y a partir de dichos estadísticos construirá un intervalo de confianza 0,95 para la media poblacional de los rendimientos de dicho híbrido en los lotes agrícolas de la región.

1. Identificar la población de referencia, la muestra y las variables aleatorias a registrar.
2. Explicar por qué las variables aleatorias a registrar son independientes y con idéntica distribución de probabilidad.
3. ¿Qué valores tendrán la esperanza y la varianza de dichas variables aleatorias?
4. Explicar por qué la media muestral y la *varianza muestral* son variables aleatorias.
5. ¿Qué es la esperanza de la media muestral? ¿Cuál será su valor?
6. ¿Qué es la varianza de la media muestral? ¿Cuál será su valor? ¿Cómo lo estimará el fitotecnista?
7. ¿Por qué es correcto afirmar que la media muestral es un estimador insesgado y consistente de la media poblacional?
8. ¿Por qué es correcto afirmar que la varianza muestral es un estimador insesgado y consistente de la varianza poblacional?
9. ¿A qué modelo se aproxima la distribución de probabilidad de la media muestral?
10. ¿De qué depende la probabilidad de que la media muestral tome un valor cercano al de la media poblacional?
11. ¿Qué riesgo corre el fitotecnista de producir un intervalo que no contenga el valor de la media poblacional? ¿Cómo hace para controlar fijar ese riesgo?

Ejercicio 3. Una serie de mediciones del contenido de hidrocarburos en los gases de escape (ppm) de 15 autos de 10 años de antigüedad matriculados en Buenos Aires tomados al azar produjo los siguientes valores:

Hidrocarburos (ppm)	381 – 261 – 552 – 207 – 729 – 295 – 511 – 425 – 414 – 382 – 403 – 523 – 391 – 575 – 599
---------------------	--

1. Identificar la población de referencia, la variable de interés y la muestra aleatoria obtenida.
2. Calcular estimaciones puntuales de la media poblacional y de la varianza poblacional de los contenidos de hidrocarburos en los gases de escape de los autos de la población de referencia.
3. Explicar qué es la media poblacional en este caso particular.
4. Dando por cierto que los valores registrados son realizaciones de variables aleatorias independientes con idéntica distribución Normal, construir un intervalo de confianza 0,99 para la media poblacional de los contenidos de hidrocarburos en los gases de escape de los autos de la población de referencia.

En relación con la precisión del intervalo de confianza 0,99 construido,

5. ¿Cómo será la precisión de un intervalo de confianza 0,95 construido con los mismos datos? Verificarlo con el cálculo apropiado.
6. ¿Cómo será la precisión de un intervalo de confianza construido con una muestra de tamaño $n = 20$? Verificarlo incorporando a la muestra los siguientes 5 nuevos registros y luego calculando un nuevo intervalo de confianza con los 20 datos.

Hidrocarburos (ppm)	476 – 510 – 556 – 276 – 648
---------------------	-----------------------------

7. ¿Cómo será la precisión de un intervalo de confianza construido con datos en los que por descuido se incluyeron 4 autos de 20 años de antigüedad?

Ejercicio 4. Una asociación de defensa del consumidor denuncia que una empresa arroceras vende paquetes etiquetados como de 1kg que en realidad pesan en promedio menos que 1 kg. Para establecer si es razonable aceptar esta denuncia, el juzgado que interviene pide el asesoramiento de peritos estadísticos que proponen obtener y examinar una muestra aleatoria de 15 paquetes de arroz entregados por la empresa en cuestión a la distribuidora. Con los pesos

de los paquetes a incluir en la muestra, se disponen a estimar el promedio y la varianza de todos los pesos de los paquetes producidos por la empresa y a realizar una prueba de hipótesis apropiada para juzgar el caso.

1. Identificar las unidades muestrales, la muestra y las variables aleatorias involucradas en este problema. ¿Por qué causas se trata de variables aleatorias independientes?
2. ¿Qué experimento aleatorio genera las variables aleatorias en cuestión?
3. Formular hipótesis verbales y estadísticas apropiadas para juzgar el caso.
4. Seleccionar un nivel de significación apropiado. Explicar qué implica el valor elegido.

Una vez obtenida la muestra y realizadas las mediciones los pesos de los paquetes resultan ser los siguientes (en gramos):

987	995	1003	945	1002	968	1002	997	1005	987	985	1007	1009	979	1003
-----	-----	------	-----	------	-----	------	-----	------	-----	-----	------	------	-----	------

5. Calcular el valor p . Explicar que mide y cómo se define.
6. Concluir la prueba de hipótesis y explicar la conclusión en términos del problema particular.
7. ¿A qué población se extiende la inferencia realizada?
8. ¿Qué error se puede haber cometido? ¿Qué consecuencias tendría? ¿De qué depende el riesgo de ese tipo de error?

Ejercicio 5. El pastoreo de las ovejas es selectivo; estos animales comen preferentemente plantas de ciertas especies vegetales y evitan otras. En la estepa Patagónica, una especie preferida por las ovejas es el pasto *Bromus pictus* y entre las especies evitadas hay un grupo de pastos duros localmente denominados coirones. Se ha postulado que los coirones podrían proteger del pastoreo a las plantas de especies preferidas que crecen cerca de ellas.

Para poner a prueba esta hipótesis, investigadores de la FAUBA deciden comparar la biomasa promedio (gramos/planta) de las plantas de *Bromus pictus* que crecen cerca de coirones con la biomasa promedio de plantas de la misma especie que crecen lejos de dichos pastos duros en un lote experimental de 5000 has sometido a pastoreo ovino en la estepa patagónica. Para ello, obtendrán, dentro del lote experimental, una muestra aleatoria de 10 plantas de *Bromus pictus* que crecen cerca de coirones y otra muestra aleatoria de 10 plantas que crecen lejos de coirones y evaluar la biomasa de cada planta.

1. Identificar las poblaciones de referencia, las muestras, las unidades muestrales y las variables aleatorias involucradas en esta investigación.
2. Explicar qué son los valores de biomasa promedio que se intenta comparar.
3. Proponer dos posibles causas materiales de la varianza de las variables aleatorias a registrar.

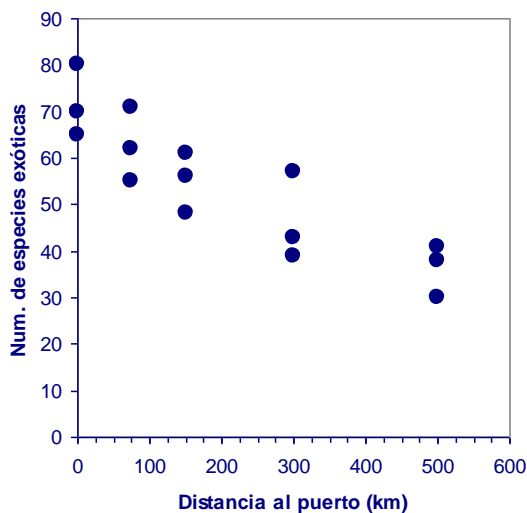
Luego de realizar el experimento propuesto, los investigadores registra los datos que figuran en siguiente la tabla:

Biomasa total [g/planta]	
Lejos de coirones	0,09 – 1,23 – 0,01 – 1,12 – 0,36 – 0,37 – 0,64 – 0,24 – 0,05 – 0,01
Cerca de coirones	0,70 – 1,84 – 0,50 – 0,89 – 0,62 – 0,85 – 1,83 – 0,79 – 0,48 – 0,51

4. Formular y poner a prueba hipótesis apropiadas para cumplir con el objetivo de los investigadores con un nivel de significación $\alpha = 0,05$.
5. Interpretar los resultados de la inferencia en términos del problema particular.
6. Explicar qué error se puede haber cometido. ¿Qué se hubiera podido hacer para que el riesgo de este tipo de error fuera menor?
7. Construir un intervalo de confianza 0,95 para la diferencia entre la biomasa promedio de las plantas de *Bromus pictus* que crecen cerca y lejos de los coirones en el lote experimental.

8. Bajo qué supuestos son válidos la prueba de hipótesis realizada y el intervalo de confianza construido.

Ejercicio 6. Las invasiones biológicas constituyen un serio problema ambiental. En una investigación sobre la invasión de la Región Pampeana por plantas exóticas (de otras regiones) se propuso que el puerto de Buenos Aires funcionaría como la principal vía de ingreso accidental de estas plantas a la región. En este caso, se esperaría que la diversidad promedio de especies de plantas exóticas fuera máxima en las cercanías de Buenos Aires y disminuyera con la distancia a este puerto. Para poner a prueba esta relación estadística se tomaron al azar 15 sitios en la Región Pampeana a distancias preestablecidas de Buenos Aires y se registró el número de especies vegetales exóticas presentes. A continuación se presenta un diagrama de dispersión construido con los datos obtenidos y parte de los cálculos necesarios para realizar un análisis de regresión lineal simple con los mismos.



$$\bar{x} = 205$$

$$\bar{y} = 54,4$$

$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = 474.000,00$$

$$\sum_{i=1}^{15} (x_i - \bar{x}) \cdot (y_i - \bar{y}) = -32.230,00$$

$$\sum_{i=1}^{15} (y_i - \bar{y})^2 = 2.849,60$$

$$\sum_{i=1}^{15} [y_i - (b_0 + b_i \cdot x_i)]^2 = 658,10$$

1. Escribir el modelo de regresión e interpretar la variable aleatoria, la variable no aleatoria y cada uno de los parámetros en términos del problema (indicar sus unidades).
2. Obtener las estimaciones puntuales de la ordenada al origen y de la pendiente de la recta de regresión. Dibujar la recta de regresión estimada.
3. Poner a prueba la siguiente hipótesis nula: *En la Región Pampeana, el número promedio de especies vegetales exóticas no varía con la distancia a Buenos Aires.*
4. Calcular e interpretar un intervalo de confianza del 95% para la pendiente de la recta de regresión.
5. Construir intervalos de confianza para los números promedio de especies vegetales exóticas a 0, 100, 200, 300, 400 y 500 km de Buenos Aires. Volcar los intervalos en el gráfico y dibujar una banda de confianza
6. Calcular e interpretar el coeficiente de determinación.

Ejercicio 7. Un ecólogo está interesado en determinar si la pubescencia de las hojas de una especie vegetal está asociada con la presencia de carbonato de calcio masivo en el suelo. Para ello, decide seleccionar al azar 100 plantas de la especie de interés dentro de un Parque Nacional y registrar si éstas crecían en suelo calcáreo o no y si tienen o no pubescencia en las hojas.

1. Identificar a las unidades muestrales, la muestra y la población de referencia.
2. ¿Cuántas variables aleatorias categóricas registra el ecólogo en cada unidad muestral?

Una vez realizado el procedimiento, los datos obtenidos por el ecólogo son los siguientes:

Frecuencias absolutas (plantas)

Tipo de suelo	Hojas	
	Pubescentes	Glabras
Calcáreo	12	22
No calcáreo	16	50

A partir de estos datos,

- Estimar y comparar las frecuencias relativas de las plantas con hojas pilosas en los suelos calcáreos y en los suelos no calcáreos.
- Explicar por qué existe incertidumbre en las estimaciones anteriores.
- ¿Qué tipo de prueba de hipótesis corresponde realizar para determinar si las diferencias entre las frecuencias relativas observadas de las plantas con hojas pilosas en los suelos calcáreos y en los suelos no calcáreos son significativas?
- Llevar a cabo la prueba hipótesis con un nivel de significación $\alpha = 0,05$.
- Escribir la conclusión de la prueba de hipótesis en términos del problema.
- Explicar qué error se puede haber cometido e indicar cómo se hubiera podido disminuir el riesgo de cometerlo.

Ejercicio 8. Si el rendimiento de un lote de maíz (t/ha) a tomar al azar en el partido de Pergamino es una variable aleatoria con distribución aproximadamente Normal con $\mu = 9,5$ y $\sigma = 0,75$.

- Identificar el experimento aleatorio que produce la variable aleatoria en cuestión.
- ¿Cuánto vale la esperanza del rendimiento de un lote de maíz tomar al azar en el partido de Pergamino? ¿Por qué nos interesa este valor?
- ¿Cuánto vale la varianza del rendimiento de un lote de maíz a tomar al azar en el partido de Pergamino?
- Nombrar dos posibles causas materiales de la varianza de los rendimientos del cultivo de maíz en el partido de Pergamino.
- ¿Cuál es el valor de rendimiento superado por el 10% de los lotes de maíz del partido de Pergamino?
- ¿Cuál es el valor del percentil 75 de la distribución de frecuencias de los rendimientos de maíz en el partido de Pergamino?
- ¿Cuánto vale la probabilidad de que el rendimiento de un lote a tomar al azar difiera de la media poblacional en menos que 2 t/ha?

Alguien que no conoce el valor de la media poblacional de los rendimientos tomará una muestra aleatoria de lotes de maíz del partido de Pergamino con el propósito de realizar construir un intervalo de confianza 0,99 para dicho valor,

- ¿Cuánto vale la probabilidad de que el intervalo a construir resulte en una estimación correcta?
- ¿De qué depende la precisión del intervalo a construir?

Ejercicio 9. Una de las fuentes de contaminación de la atmósfera con dióxido de carbono es la respiración de los suelos agrícolas. La cantidad de CO_2 liberado por del suelo a la atmósfera depende de factores tales como la temperatura, humedad, fertilidad y aireación del suelo, del tipo y cantidad de materia orgánica que contiene y de la composición y abundancia de su meso y microbiota. Estos factores son, a su vez, afectados por las diferentes prácticas de cultivo. Con el fin de evaluar el posible efecto del sistema de labranza sobre la respiración del suelo, un equipo integrado por Ingenieros Agrónomos y Licenciados en Ciencias Ambientales se propone comparar la respiración promedio del suelo de lotes agrícolas del partido de Rojas cultivados con labranza mecánica convencional y con siembra directa. Para ello, seleccionará

al azar 10 lotes agrícolas dentro del partido, dividirá cada lote en dos, en una mitad cultivará soja con labranza convencional y en la otra mitad cultivará soja con siembra directa. Durante el tercer mes de cultivo, procederá a medir la respiración del suelo en cada mitad.

1. Identificar las unidades muestrales, la muestra y las variables aleatorias involucradas en este procedimiento.
2. Escribir las hipótesis verbales de la comparación propuesta y formular las hipótesis estadísticas correspondientes. Explicar a qué corresponden los parámetros incluidos en las hipótesis estadísticas.

Luego de realizado el experimento, los datos obtenidos son los siguientes:

Respiración del suelo [mg CO₂ liberado / día /100 g de suelo]

Lote	1	2	3	4	5	6	7	8	9	10
Lab. Conv.	82,5	96,8	99,7	86,2	95,0	96,1	88,0	89,8	91,7	94,6
Siembra directa	82,1	89,1	87,3	78,0	91,9	97,4	81,3	78,7	94,9	92,3

Dando por cierto que la diferencia entre los niveles de respiración del suelo de un lote a tomar al azar bajo labranza convencional y bajo siembra directa es una variable con distribución Normal,

3. Calcular el valor p para poner a prueba la hipótesis nula que con nivel de significación de significación $\alpha = 0,01$.
4. Construir e interpretar un intervalo de confianza del 99% para la diferencia de respiración promedio entre lotes cultivados con soja con labranza convencional y con siembra directa en el partido de Rojas.
5. ¿El intervalo de confianza del 99% construido es consistente con el resultado de la prueba de hipótesis? Explicar en qué sentido.
6. ¿Qué tipo de error que se puede haber cometido en esta prueba? ¿Qué significado tendría en este caso? ¿De qué dependía su probabilidad?
7. ¿A qué población se extiende la inferencia realizada?

Ejercicio 10. Un problema ambiental derivado del uso incorrecto de fertilizantes en la agricultura es la contaminación de las aguas subterráneas con nitratos que se lixivian del suelo antes de que las plantas puedan absorberlos. El agua con 10 o más ppm de nitratos es tóxica, especialmente para los niños, porque inhibe la hemoglobina. Para evaluar la severidad de este problema en una localidad agrícola, un grupo de técnicos se dispone a obtener una muestra aleatoria de agua de la napa freática en 15 puntos a tomar al azar dentro de la localidad. Con los datos a obtener realizarán una prueba de hipótesis para decidir si se puede aceptar que el contenido promedio de nitratos en el agua de la localidad está por debajo del límite de toxicidad para los niños.

1. Identificar la población y las variables aleatorias en cuestión.
2. Formular hipótesis verbales y estadísticas apropiadas para la prueba propuesta. Explicar qué es el parámetro al que se refieren las hipótesis estadísticas.
3. Elegir un nivel de significación ¿Cuál es el valor crítico correspondiente?

Una vez obtenida la muestra, los datos registrados son los siguientes:

Contenido de nitratos [ppm]	6,1 - 2,4 - 16,6 - 18,3 - 10,2 - 14,9 - 12,0 - 2,0 - 3,4 - 5,0 - 18,3 - 2,4 - 3,5 - 8,6 - 4,1
------------------------------------	---

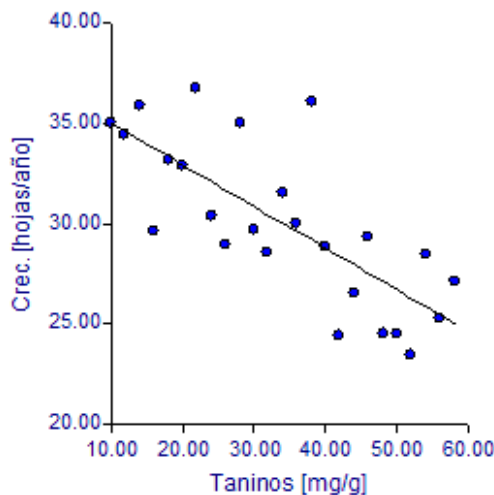
4. Explicar a qué nos referimos cuando decimos que los datos presentados constituyen *realizaciones* de las variables aleatorias en cuestión.

Suponiendo que estas variables tienen distribución Normal,

5. ¿Cuál es el estadístico de prueba apropiado para poner a prueba estas hipótesis?

- Completar la prueba de hipótesis y concluir en términos del problema de la contaminación del agua de las napas con nitratos.
- ¿Qué error se puede haber cometido? ¿Qué significado tiene en relación con la evaluación de la severidad del problema de contaminación del agua subterránea? ¿Cómo se debió haber hecho el experimento para poder asegurar razonablemente que el riesgo de este tipo de error es muy bajo?

Ejercicio 11. A continuación se presenta parte de los resultados de un análisis de regresión lineal simple realizado para examinar la relación estadística entre el crecimiento de arbolitos de *Cecropia peltata* [hojas nuevas/año] y el contenido de taninos en sus tejidos [mg taninos/g de tejido].



$$n = 25$$

$$b_0 = 37,06 \text{ hoja/año}$$

$$s_{\hat{\beta}_0} = 1,38 \text{ hoja/año}$$

$$b_1 = -0,21 \text{ (hoja/año)} / \left(\frac{\text{mg}}{\text{g}}\right)$$

$$s_{\hat{\beta}_1} = 0,04 \text{ (hoja/año)} / \left(\frac{\text{mg}}{\text{g}}\right)$$

$$R^2 = 0,57$$

- Escribir el modelo de regresión lineal completo correspondiente al análisis presentado y explicar el significado que cada término tiene en este caso. Identificar los parámetros, la/s variable/s aleatoria/s y la variable no aleatoria.
- Enunciar en palabras la relación estadística propuesta por el modelo.
- Construir un intervalo del 95% de confianza para la ordenada al origen de la recta de regresión. ¿Cómo se interpreta este parámetro en este caso?
- Poner a prueba la siguiente hipótesis nula: *la pendiente de la recta de regresión es igual a cero* ($\alpha = 0,05$).
- A partir de este análisis ¿Qué se puede decir de la relación estadística entre el contenido de taninos y el crecimiento de los arbolitos de *Cecropia peltata*?
- Explicar qué implica el nivel de significación $\alpha = 0,05$ en este caso.
- Interpretar el coeficiente de determinación R^2 en términos de este problema.
- ¿Qué condiciones (requisitos y supuestos) deben ser satisfechas para que la inferencia realizada sea válida (esto es para que el nivel de significación sea verdaderamente el seleccionado)?

Ejercicio 12. 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.

- ¿En qué consiste el procedimiento que el técnico debe aplicar para obtener la muestra aleatoria? ¿Qué variables aleatorias registrará?
- Explicar por qué causas la media muestral de los contenidos de sales del agua de los 25 pozos que se incluyen en la muestra es una variable aleatoria.

- ¿Qué es la esperanza de la media muestral en este caso particular? ¿Qué valor tiene?
- ¿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?

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,

- ¿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?
- ¿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?

Ejercicio 13. Una empresa agropecuaria evalúa las alternativas de sembrar trigo en dos localidades A y B. El rendimiento mínimo necesario para obtener un resultado económico positivo es de 38 quintales/ha. La empresa cuenta con los siguientes datos de rendimiento de cultivos de trigo escogidos al azar en cada localidad ([quintales/ha]):

Rendimiento de trigo [qq/ha]	
Localidad A	35 - 42 - 41 - 41 - 37 - 41 - 40 - 43 - 40 - 38 - 41 - 39 - 39 - 42 - 38
Localidad B	38 - 36 - 39 - 41 - 40 - 35 - 36 - 36 - 40 - 39 - 35 - 38 - 42 - 39 - 39

- Identificar a las unidades muestrales, las muestras y las poblaciones.
- Dando por cierto que es razonable suponer que los rendimientos de trigo en estas localidades se ajustan razonablemente a distribuciones y que las varianzas de los rendimientos de trigo tienen igual valor en ambas localidades, completar la prueba de hipótesis para determinar si la diferencia entre las estimaciones de las medias poblacionales es significativa con $\alpha = 0,05$.
- Construir un intervalo de confianza 0,95 para la diferencia entre las medias poblacionales de los rendimientos de trigo de ambas localidades (*ver ayuda en ejercicio 6.12*).
- Comunicar las conclusiones en la forma de una recomendación para la empresa.
- Identificar qué errores se pudo haber sido cometido y explicar qué significado tendrían para la empresa.

Ejercicio 14. Un ingeniero agrónomo necesita una estimación precisa y confiable de la cantidad de pasto presente en un lote de 32 has sembrado con una pastura.

- Detallar los pasos del procedimiento que el ingeniero deberá seguir para obtener la estimación que necesita. Cuando corresponda, presentar las fórmulas que aplicará.
- Explicar por qué el intervalo de confianza a construir es un intervalo aleatorio. ¿Qué consecuencia tiene su carácter aleatorio?
- ¿Cuál será la medida de la confianza a depositar en el intervalo a construir? ¿Cómo la controla el ingeniero?
- ¿Cómo controla el ingeniero la precisión del intervalo a obtener? ¿Qué información necesita para ello?

Ejercicio 15. Los frutos del curupí (*Sapium haematospermum*), un árbol nativo de la Región Chaco-Pampeana, están divididos en tres compartimentos (lóculos). Cada uno de estos compartimentos contiene un óvulo que produce una semilla si es fecundado o que aborta si no es fecundado. En consecuencia, cada fruto de curupí contiene entre 0 y 3 semillas. Un botánico que investiga la biología reproductiva de esta especie intenta determinar si los insectos influyen en la polinización de sus flores. Para ello, decide seleccionar dentro de una Reserva de Vida Silvestre dos muestras aleatorias de 100 flores de curupí y cubrir las flores de la primera muestra con bolsitas de malla que impiden el acceso de los insectos. Luego de formados los frutos, los cosechará, los clasificará por el número de semillas que contienen y registrará las frecuencias de frutos con diferente número de semillas en cada muestra.

1. Identificar la población de referencia, las muestras, los tratamientos y las variables aleatorias involucradas en esta investigación.
2. Formular hipótesis estadísticas apropiadas para decidir con un nivel de significación $\alpha = 0,05$ si los datos a obtener permitirán aceptar que los insectos afectan la fecundación de los óvulos de las flores de curupí.

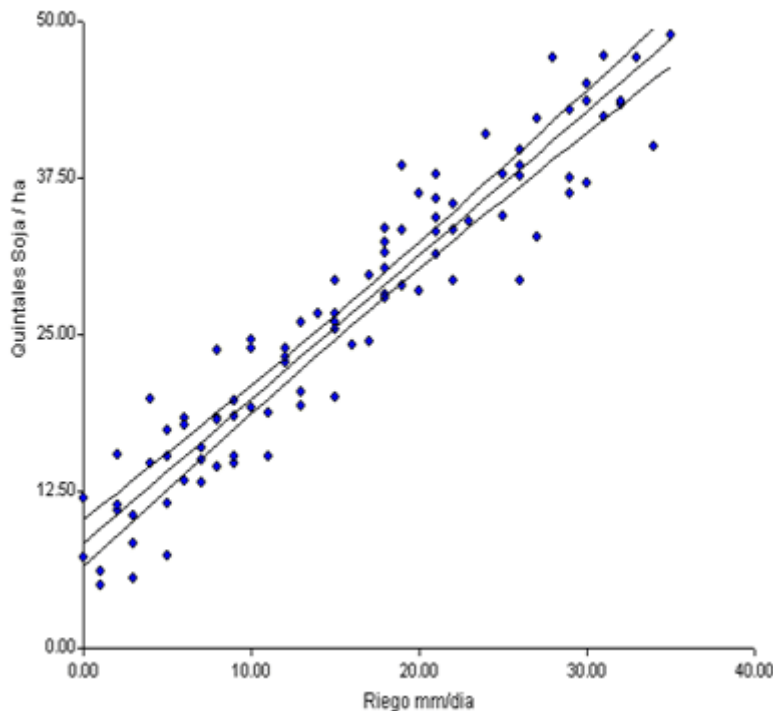
Una vez completado el experimento, el biólogo cuenta con los datos que se muestran en la siguiente tabla:

Frecuencias absolutas observadas

Números de semillas	0	1	2	3
Con bolsita	7	24	43	26
Sin bolsita	3	11	51	35

3. Completar la prueba de hipótesis y escribir la conclusión que el biólogo debería presentar.
4. ¿A qué población se extiende la inferencia realizada?

Ejercicio 16. En áreas donde las lluvias no alcanzan para cubrir las necesidades de agua de los cultivos, se recurre al uso de equipos de riego para completar el suministro de agua para las plantas. La operación de estos equipos produce costos que, para ser justificados, deben ser compensados por el valor del incremento de rendimiento que producen. En una localidad agrícola, se obtuvo un conjunto de datos de los rendimientos de una muestra de cultivos de soja (quintales/ha) tratados con diferentes intensidades de riego (mm/día) establecidas de antemano. Con estos datos, se realizó el análisis de regresión lineal simple parte de cuyos resultados se transcribe a continuación.



	EE	l (95%)	u (95%)	valor p
$b_0 = 8,39$ qq/ha	0,72	6,97	9,81	<0,0001
$b_1 = 1,15$ (qq/ha)/ (mm/día)	0,04	1,07	1,22	<0,0001

1. Explicar qué tipo de relación es razonable proponer que existe entre el rendimiento [quintales/ha] y la intensidad de riego [mm/día].

2. Escribir el modelo de regresión propuesto y explicar cada parámetro en términos de este problema.
3. Explicar qué son los valores que están en la columna de la tabla de resultados encabezada con EE.
4. Presentar un intervalo del 95% de confianza para el rendimiento promedio de los cultivos de soja de esta localidad que no reciben riego.
5. Presentar una estimación puntual del rendimiento promedio de los cultivos de soja de esta localidad que reciben 30 mm de agua de riego por día.
6. Presentar una estimación puntual del cambio en el rendimiento promedio asociado con un aumento de 10 mm/día en la intensidad de riego.
7. Explicar por qué las estimaciones presentadas son realizaciones de variables aleatorias. ¿Qué distribución de probabilidad se supone que tiene cada una de esas dos variables aleatorias?
8. Sobre la base de los resultados del análisis, ¿puede concluirse, con un nivel de significación $\alpha=0,01$, que en la localidad en cuestión el rendimiento promedio de los cultivos de soja aumenta con la intensidad de riego? Justificar la respuesta.
9. ¿Qué tipo de error se puede haber cometido en la respuesta a la pregunta anterior? Explicar qué consecuencias tendría dicho error. ¿Qué riesgo hemos corrido de cometerlo?

Ejercicio 17. Un productor avícola está interesado en evaluar el efecto de la administración de un suplemento dietario sobre el crecimiento de los pollos. Para ello decide tomar de su establecimiento dos muestras aleatorias de 61 pollos de 200 g cada una y alimentar a los pollos de la primera muestra con la ración habitual y a los de la segunda con la ración habitual más el suplemento en cuestión. Luego de 4 semanas de pesará los pollos de ambas muestras y utilizará los datos estimar la diferencia en el crecimiento promedio de los pollos sometidos a los dos tratamientos tratamiento.

1. Identificar las unidades muestrales, las muestras, la población, los tratamientos y las variables aleatorias en cuestión.
2. Explicar a qué parámetros se refiere la estimación a realizar.

Una vez realizado el experimento y obtenidos los datos, el productor avícola calcula los estadísticos que se muestran en la tabla.

Tratamiento	\bar{x} [g]	s [g]
Ración habitual	734	40,4
Ración habitual + suplemento	766	45,5

3. Dando por cierto que es razonable suponer que las varianzas de los pesos de los pollos no difieren entre estos tratamientos, construir un intervalo de confianza 0,99 para la diferencia entre las medias poblacionales de los aumentos de peso de pollos alimentados durante 4 semanas con ración habitual y con ración habitual más suplemento dietario (*ver ayuda en ejercicio 6.12*).
4. Interpretar el intervalo de confianza construido en términos del interés del productor avícola.
5. Explicar qué error puede haberse cometido en esta inferencia. ¿Qué nivel de confianza se puede tener en no haberlo cometido y por qué?

Ejercicio 18. El los prospectos publicados por una compañía productora de semillas se asegura que, en la Pampa Ondulada, el rendimiento promedio de uno de sus híbridos de maíz supera las 10 t/ha. Para decidir si se puede aceptar esta afirmación, los productores de un grupo CREA decide poner a prueba este híbrido cultivándolo en 12 parcelas de 1 ha seleccionadas al azar en los establecimientos del grupo y registrando los rendimientos a obtener.

1. Identificar la población y las variables aleatorias en cuestión.

2. Formular hipótesis verbales y estadísticas apropiadas para la prueba propuesta. Explicar qué es el parámetro al que se refieren las hipótesis estadísticas.
3. ¿Cuál es el estadístico de prueba apropiado para poner a prueba estas hipótesis? ¿Qué supuesto implica su uso?
4. Elegir un nivel de significación y explicar qué implica.
5. ¿Cuál es el valor crítico correspondiente al tamaño de muestra y nivel de significación elegidos?

Una vez cosechadas las parcelas de prueba los rendimientos registrados son los que figuran en la tabla.

Parcela	1	2	3	4	5	6	7	8	9	10	11	12
Rendimiento (t/ha)	12,4	11,0	10,5	11,7	9,9	12,0	8,9	9,7	11,5	9,2	10,3	11,5

6. Completar la prueba de hipótesis y concluir en términos del objetivo de los productores del grupo CREA.
7. ¿Qué error se puede haber cometido? ¿Qué consecuencia tendría?

Ejercicio 19. En los últimos años, el hábitat de la fauna pampeana fue modificado porque gran parte de los pastizales fueron reemplazados por cultivos. Para evaluar la influencia de dichos cambios sobre la composición de la avifauna, un grupo de ornitólogos decide obtener una muestra aleatoria de 1000 aves avistadas en el partido de Navarro y clasificarlas según su hábito alimentario en rapaces, carroñeras, granívoras y otras. El objetivo de los ornitólogos es determinar si, en el área de estudio, las frecuencias relativas de los grupos mencionados difieren de los valores históricamente reportados para la avifauna pampeana: rapaces (0,10), carroñeras (0,13), granívoras (0,58) y otras (0,19).

1. Identificar el tipo de prueba de hipótesis requerido y formular las hipótesis estadísticas apropiadas para la investigación propuesta.

Una vez obtenida la muestra, entre las 1000 aves avistadas había 115 rapaces, 116 carroñeras, 605 granívoras y 164 de otros hábitos alimentarios.

2. Completar la prueba de hipótesis con un nivel de significación $\alpha = 0,05$ y escribir la conclusión en términos del objetivo de los ornitólogos.
3. Controlar el cumplimiento de los requisitos para la validez de la prueba.
4. ¿Qué representa, en este caso particular, el nivel de significación $\alpha = 0,05$ elegido?