Estadística (Química)

Práctica 5 - Estimación - Intervalos de Confianza

Comentario: En todos los ejercicios propuestos

a) defina las variables aleatorias y los parámetros involucrados.

b) de ser posible indique:

   i. la distribución de las variables aleatorias

   ii. el significado intuitivo de los parámetros.

c) compare los resultados de hacer las cuentas a mano con las salidas obtenidas con el R, de manera de chequear las primeras y aprender a usar las segundas, en aquellos ejercicios en los que ambas cosas sean posibles.

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2.

   (a) Probar que $\overline{X}_n^2$ no es un estimador insesgado de μ^2.

   (b) ¿Para qué valores de k es $\widehat{\mu}^2 = \left(\overline{X}_n^2 - ks^2\right)$ un estimador insesgado de μ^2?

2. Sea $X_1, \ldots . X_n$ una muestra aleatoria de una distribución Bernoulli de parámetro p y sea $T_n = \sum_{i=1}^{n} X_i$. Consideremos un nuevo parámetro $\theta = p\,(1-p)$. Mostrar que

$$\frac{T_n\,(n - T_n)}{n\,(n-1)}$$

   es un estimador insesgado de θ.

3. El volumen de una burbuja se calcula midiendo su diámetro y usando la relación

$$V_0 = \frac{\pi}{6} D_0^3$$

   donde V_0 y D_0 son el volumen y dimetros verdaderos de la burbuja. Se dispone de $n = 5$ mediciones independientes del dimetro (en milmetros) de la burbuja. La media muestral resulta ser $\bar{D} = 2.3$. Cada medicin del dimetro tiene distribucin $\mathcal{N}(D_0, 0.01)$.

   (a) Estime el volumen del cilindro y aproxime el desvo estndar del estimador.

   (b) Si no conociramos el verdadero desvo estndar de cada medicin del dimetro, pero nos dijeran que el desvo estndar muestral calculado con las 5 observaciones fue $S = 0.02$. Cmo podramos estimar el desvo estndar del estimador del volumen propuesto en (a)?

4. Para determinar el volumen de un cilindro se utiliza la formula $V_0 = \pi R_0^2 A_0$ donde V_0, R_0 y A_0 son el volumen, radio de la base y altura verdaderas del cilindro. Cada medicin del radio y de la altura siguen distribuciones $\mathcal{N}(R_0, 0.1)$ y $\mathcal{N}(A_0, 0.05)$ respectivamente. De 5 mediciones del radio y 3 de la altura se obtuvieron las medias muestrales $\bar{R} = 1$ y $\bar{A} = 3$. Asumamos que todas las mediciones son independientes.

   (a) Estime el volumen del cilindro y el desvo estndar del estimador propuesto.

   (b) Si el metrlogo quisiera mejorar la precisin de su estimacin pero slo dispone del tiempo para volver a medir slo una vez el radio la altura. Cul le recomendara que repita? Por qu?

5. Para determinar la constante de un resorte, se puede medir la oscilacin de una masa m_0 fijada en uno de sus extremos. Mediante la frmula $k_0 = 4\pi^2 m_0/T_0^2$ donde k_0 es la constante del resorte y T_0 es el perodo de oscilacin de dicha masa. Las cantidades m_0, k_0 y T_0 son los valores verdaderos. Se realizan mediciones de la masa y del perodo. Asumamos que cada medicin de la masa y del perodo siguen distribuciones $\mathcal{N}(m_0, 0.1)$ y $\mathcal{N}(T_0, 0.05)$ respectivamente. De 5 mediciones de la masa y 4 del perodo se obtuvieron medias muestrales $\bar{m} = 0.52$ y $\bar{T} = 1.3$. Asumiendo que todas las mediciones son independientes, estime la constante del resorte y el desvo estandar de la estimacin.

6. Se registró el valor (en Kg) de la reducción del peso, de cada uno de 16 pacientes elegidos al azar, después de una semana de tratamiento. El promedio de esos 16 valores fue de 3.42Kg. Suponga que la pérdida de peso luego de una semana de tratamiento es una variable aleatoria con distribución normal.

   (a) Construya un intervalo de confianza del 99% para el valor medio poblacional de la reducción del peso después de una semana de tratamiento.

      i. en el caso σ conocida con $\sigma = 0.68$Kg.

      ii. en el caso σ desconocida y $s = 0.68$Kg.

   (b) Compare la longitud de los intervalos obtenidos.

7. Diez mediciones de recuperación de bromuro potásico por cromatografía de gas-líquido en muestras de tomates de cierta partida arrojaron una media muestral de $782\mu g/g$ y un desvío $s = 16.2\mu g/g$. Suponga que las mediciones tienen distribución normal.

   (a) Halle un intervalo de confianza del 95% para la media (μ) de las mediciones en esta partida de tomates.

   (b) Suponiendo que el error de medición debido al método es despreciable respecto a la variabilidad entre los tomates y que las 10 mediciones se realizan para cada uno de 10 tomates, ¿cómo debe interpretarse la media μ? ¿Cómo debe interpretarse σ^2?

   (c) En cambio, si las 10 mediciones se realizaron sobre el mismo tomate, ¿cómo puede interpretarse μ? ¿Cómo debe interpretarse σ^2?

   (d) Halle un intervalo de confianza del 95% para la varianza $\left(\sigma^2\right)$ de las mediciones.

   (e) Halle un intervalo de confianza del 95% para la desviación estándar (σ) de las mediciones.

8. Se hicieron varias mediciones del contenido de glucosa de una solución. Suponga que estas mediciones siguen un modelo de Gauss sin sesgo, es decir, el modelo descripto en el ejercicio 3 de la práctica 3, cuando los errores tienen distribución normal.

   (a) Escriba el modelo. Se calculó el intervalo de confianza del 95% para la media, que resultó ser $(10.28, 11.32)$. ¿Qué significa "la media" en este problema?

   (b) Decir si es verdadero o falso, y explicar:

      i. Un 95% de las mediciones caerán en ese intervalo.

      ii. Hay una probabilidad de 0.95 de que la próxima medición caiga en el intervalo.

      iii. Alrededor del 95% de las veces que uno realice el ensayo y construya el intervalo de confianza, éste contendrá la verdadera concentración de glucosa de la solución.

      iv. La probabilidad de que el intervalo $(10.28, 11.32)$ contenga a la verdadera concentración de glucosa es de 0.95.

9. A partir de un gran número de mediciones, se sabe que un método para determinar la cantidad de manganeso en un mineral comete errores aleatorios con distribución normal de media cero y desviación estándar 0.09.

(a) Se hicieron 5 mediciones de un mismo mineral y se obtuvo un valor promedio de 7.54, calcule un intervalo de confianza con nivel del 99% para la cantidad de manganeso verdadera que contiene ese mineral.

(b) Ídem a), pero si se hicieron 10 mediciones.

(c) ¿Cuántas mediciones habría que hacer para que el intervalo de confianza al 99% tenga una longitud ≤ 0.10?

10. Considere el ejercicio 4 de la Práctica de Estadística Descriptiva.

(a) Usando las medias y desviaciones estándar calculadas con el R, calcule un IC del 95% para la media de las mediciones hechas por cada uno de los grupos. ¿Los intervalos obtenidos tienen nivel de confianza exacto o aproximado?

(b) Observe que el paquete estadístico también puede calcular automáticamente los ICs del inciso anterior. Considere la instrucción:

```
t.test(GRUPO1,alternative= "two.sided")
```

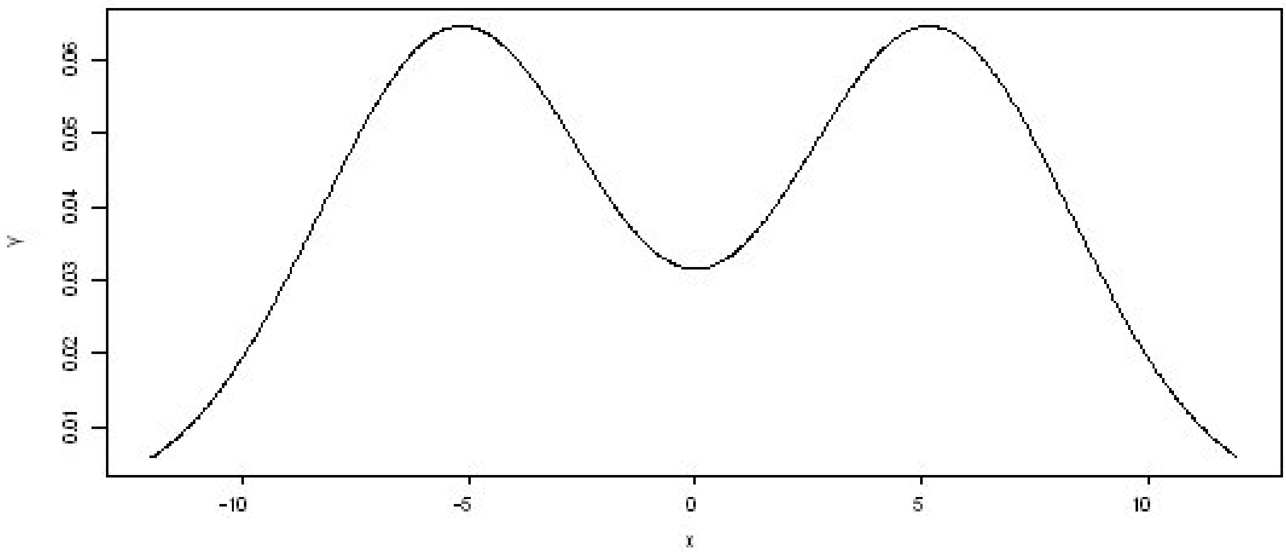
(c) ¿Es razonable pensar que ambos grupos de estudiantes están midiendo la concentración de ion nitrato sin sesgo? En caso contrario, ¿puede saberse si uno solo o ambos grupos hacen mediciones sesgadas?

11. Ciudad Gótica tiene numerosas casas alquiladas. Para realizar un estudio se eligen 400 al azar y se averigua el alquiler que pagaron sus ocupantes el mes anterior, resultando un promedio muestral $\overline{x} = \$184$ y un desvío estándar muestral $s = \$80$. Se dibuja un histograma con los alquileres registrados en la muestra y se ve que no sigue la curva normal.

(a) Si es posible, halle un intervalo de confianza del 68% aproximadamente para el alquiler medio de las casas de alquiler ocupadas en Ciudad Gótica. Si no es posible, justifique.

(b) Indique si es verdadero o falso: Para alrededor del 68% del total de casas de alquiler ocupadas en la ciudad, el alquiler fue de entre \$180 y \$188.

12. Una balanza comete errores aleatorios que siguen la densidad que se muestra en el gráfico:



La esperanza de esta densidad es 0 microgramos y su desviación estándar es 6 microgramos. Se hacen 4 mediciones de un peso y se requiere un intervalo de confianza (exacto o aproximado) para el peso verdadero.

(a) Para hallar el intervalo de confianza, ¿puede usarse la curva normal? ¿Y la t de Student? Justifique.

(b) Ídem con 100 mediciones. Dé el intervalo correspondiente para un nivel del 95% e indique si es exacto o aproximado.

(c) Se hacen 100 mediciones, en 57 de las cuales se comete un error positivo. Hallar un intervalo de confianza de nivel 95% aproximadamente para la proporción de mediciones en las que se cometen errores positivos. Investigar las siguientes instrucciones de R, y compararlas con la resolución a mano del ejercicio.

```
prop.test(x = 57,n = 100,correct=TRUE)
prop.test(x = 57,n = 100,correct=FALSE)
binom.test(x = 57,n = 100)
```

13. Para un estudio de mercado 277 personas degustan un nuevo licor y 69 de ellas desaprueban el nuevo sabor. Construya un intervalo de aproximadamente 95% de confianza para la verdadera proporción poblacional p de personas que aprueban el nuevo licor.

14. Un fabricante asegura a una compañía (que le compra un producto en forma regular) que el porcentaje de productos defectuosos no es mayor del 5%. La compañía decide comprobar la afirmación del comerciante seleccionando al azar de su inventario 200 unidades de este producto y probándolas. En la muestra encuentran 19 unidades defectuosas.

(a) Construya un intervalo del 95% aproximadamente de confianza para la verdadera proporción de unidades defectuosas.

(b) ¿Tiene razones la compañía para sospechar de la afirmación del fabricante? Justifique.

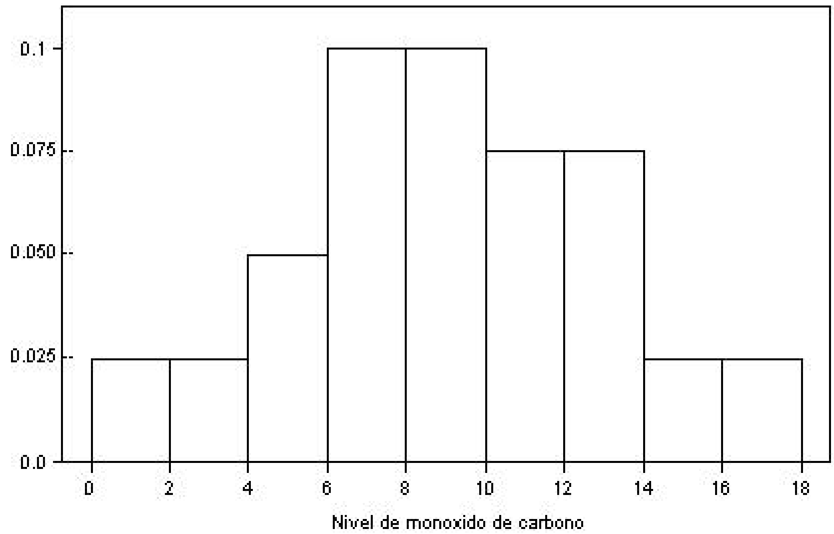
15. Se desea construir un intervalo de confianza para la media μ del puntaje de cierta prueba de lectura diseñada para los alumnos de tercer grado de la ciudad de Buenos Aires. Supongamos que, de estudios anteriores, se puede inferir que el desvío estándar poblacional es $\sigma = 12$.

(a) Se extrae una muestra al azar de 100 niños de tercer grado de la ciudad de Buenos Aires, a los que se les toma la prueba. Calcule la longitud del intervalo de confianza de nivel 95% basado en dicha muestra. El intervalo de confianza, ¿es exacto o aproximado? ¿Que hipótesis deben satisfacer las variables involucradas?

(b) Ídem (a) pero si el presupuesto permitiese sólo evaluar a 10 niños.

(c) Halle el menor valor de n para el cual se satisface el requerimiento de los investigadores de obtener un intervalo del 95% de confianza con una longitud de a lo sumo 2. Especifique los supuestos utilizados.

16. 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 proporción de esquinas que presentan 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 aleatoriamente elegidas un viernes a la tarde. Los datos obtenidos se representan en el siguiente histograma de densidad.



(a) En base a ellos, la estimación de p resulta ser ....

(b) Para completar el análisis de los datos del ítem anterior se decide calcular un intervalo de confianza de nivel aproximado 0.95 para p. Calcularlo, y dar los extremos numéricos de dicho intervalo: (...................,..........................).

17. Considere el ejercicio 6 de la Práctica de Estadística Descriptiva. Trabajaremos con las nubes Tratadas.

(a) ¿Puede calcular un intervalo de confianza de nivel 0.95 (exacto o aproximado) para la cantidad esperada de agua caída de una nube tratada con un bombardeo de átomos? Para responder este ítem, recuerde lo realizado al analizar los datos en la práctica de descriptiva.

(b) ¿Puede calcular un intervalo de confianza de nivel 0.95 para el valor esperado del logaritmo de la cantidad de agua caída de una nube tratada con un bombardeo de átomos? Resuelva usando el R. A partir del intervalo calculado, ¿puede ahora responder positivamente al ítem (a)?