

ESTADÍSTICA APLICADA
Primer Curso del Grado en Bioquímica (2018-2019)

Tema 3: ESTIMACIÓN PUNTUAL

3.1. Se planea tomar una muestra de 10 cerdos enanos (*mini-pigs*) macho y medir el nivel de creatinina sérica (en mg/dl) en cada uno de ellos. Suponemos que el nivel X de creatinina en uno de estos animales elegido al azar sigue una distribución $N(\mu = 1.5, \sigma = 0.5)$. ¿Cuál es la distribución de la media muestral? ¿Cuál es la probabilidad de que la media muestral sea superior a 1.4 mg/dl?

3.2. Una epidemióloga está planificando un estudio sobre la incidencia de cierto trastorno endocrino en la población de mujeres de un país. Planea escoger una muestra aleatoria de n mujeres y utilizar la proporción muestral \hat{p} de las que padecen dicho trastorno como estimador de la proporción poblacional p . Supongamos que en realidad $p = 0.12$.

- a) Determina la probabilidad de que el *error absoluto*, $|\hat{p} - p|$, cometido al estimar p mediante \hat{p} sea menor que 0.03 si $n = 100$.
- b) Determina la probabilidad de que el error absoluto cometido al estimar p mediante \hat{p} sea menor que 0.03 si $n = 200$.
- c) Determina el mínimo tamaño muestral n necesario para que el error absoluto sea menor que 0.01 con una probabilidad del 90 %.

3.3. Consideremos los datos de la relación 1 de problemas sobre distrofia muscular de Duchenne. Representa gráficamente el histograma de los niveles de piruvato quinasa (PQ) en el grupo de control (no portadoras) para comprobar que tiene aproximadamente una forma gaussiana. Suponiendo que las medidas de PQ en dicho grupo siguen una distribución $N(\mu, \sigma)$, obtener los estimadores de máxima verosimilitud (emv) de μ , σ^2 y σ .

Si se midiera el nivel de PQ en una nueva muestra aleatoria de 5 mujeres no portadoras, ¿qué distribución seguiría la media muestral?.

¿Cómo se explica que un modelo paramétrico (como el gaussiano) que podría generar observaciones negativas sea un modelo adecuado para una variable como la PQ, que sólo puede tomar valores no negativos?

3.4. En el fichero **neuronspikes.txt** encontramos los tiempos (en ms) entre potenciales de acción (impulsos eléctricos) consecutivos de una determinada neurona. Supongamos que dicha variable sigue una distribución exponencial de parámetro λ . Representa gráficamente un histograma de los datos. Halla el emv de λ . ¿Cuál es la distribución de la v.a. “Número de potenciales de acción que tienen lugar en un ms”? ¿Cómo estimarías el parámetro de esta distribución?

3.5. La concha del caracol de tierra *Limocolaria martesiana* presenta dos posibles aspectos: a rayas o pálida. Se extrae una muestra de este tipo de caracoles y se observa que hay 26 con la concha a rayas y 17 con la concha pálida. ¿Cuál es la emv de la proporción de caracoles con la concha a rayas? ¿Cuál sería aproximadamente la distribución de probabilidad de la proporción de conchas rayadas en una muestra de 50 caracoles de tierra? ¿Cuál es la distribución de probabilidad del número de conchas rayadas en esa muestra de 50 caracoles?

3.6. Una pipeta aforada de clase A de 25 ml vierte un volumen medio (o esperado) de 25.00 ml con una desviación típica de 0.03 ml.

- a) Se realizan 30 vertidos independientes con la pipeta. ¿Qué distribución sigue el promedio de los 30 vertidos? ¿Qué probabilidad hay de que el promedio esté entre los valores 25 ± 0.01 ml?
- b) ¿Cuántos vertidos habría que realizar para que el error del vertido promedio fuese, en valor absoluto, inferior a 0.01 con una probabilidad 0.99?

3.7. La presión parcial de oxígeno, PaO_2 , es una medida de la cantidad de oxígeno en la sangre. Supongamos que la distribución de los niveles de PaO_2 en recién nacidos tiene una media de 38 mmHg y una desviación típica de 9. Si se toma una muestra de 35 recién nacidos, ...

- a) ... ¿cuál es la distribución de probabilidad de la media muestral \bar{X} ?
- b) ... ¿cuál es la probabilidad de que la media muestral sea mayor que 37?
- c) ... ¿cuál es la probabilidad de que la media muestral sea mayor que 40?

3.8. El conjunto de datos `168Perros.txt` contiene medidas de los niveles de glucosa (en mg/dl) y alanina aminotransferasa (ALT, U/l) en una muestra de 168 perros aparentemente sanos¹.

- a) Dibuja un histograma de la variable ALT e indica razonadamente si los datos se podrían modelizar mediante una distribución normal. Dibuja después un histograma del logaritmo de dicha variable y responde a la misma pregunta.
- b) Si quisieramos ajustar un modelo $N(\mu, \sigma)$ a $\log(\text{ALT})$, ¿cómo estimaríamos los valores de los parámetros μ y σ ? Dibuja la densidad normal con los parámetros estimados sobre el histograma para comprobar la bondad del ajuste.
- c) Bajo la hipótesis de normalidad de (b), ¿qué probabilidad hay de que un perro sano tenga un nivel de $\log(\text{ALT})$ superior a 4?. Compara esta probabilidad con la frecuencia observada en los datos.

¹Fuente de los datos: Kaneko, Harvey y Bruss (2008). *Clinical Biochemistry of Domestic Animals*. Elsevier Academic Press.