Homework 1 - Statistical Inference

1 Usando la identidad

$$(\hat{\theta} - \theta) = [\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta] = [\hat{\theta} - E(\hat{\theta})] + B(\hat{\theta}),$$

demuestre que

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + (B(\hat{\theta}))^2.$$

2 **a** Si $\hat{\theta}$ es un estimador insesgado para θ, ¿cuál es $B(\hat{\theta})$?

b Si $B(\hat{\theta}) = 5$, ¿cuál es $E(\hat{\theta})$?

3 Suponga que $\hat{\theta}$ es un estimador para un parámetro θ y $E(\hat{\theta}) = a\theta + b$ para algunas constantes diferentes de cero a y b.

a En términos de a, b, y θ, ¿cuál es $B(\hat{\theta})$?

b Encuentre una función de $\hat{\theta}$, por ejemplo $\hat{\theta}$,* que es un estimador insesgado para θ.

4 Consulte los Ejercicios 1 y considere el estimador insesgado $\hat{\theta}^*$ que usted propuso en el Ejercicio 3.

a Exprese MSE $(\hat{\theta}^*)$ como función de $V(\hat{\theta})$.

b Dé un ejemplo de un valor para a para el cual $\text{MSE}(\hat{\theta}^*) < \text{MSE}(\hat{\theta})$.

c Dé un ejemplo de valores para a y b para los cuales $\text{MSE}(\hat{\theta}^*) > \text{MSE}(\hat{\theta})$.

5 Suponga que $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $V(\hat{\theta}_1) = \sigma_1^2$, y $V(\hat{\theta}_2) = \sigma_2^2$. Considere el estimador $\hat{\theta}_3 = a\hat{\theta}_1 + (1 - a)\hat{\theta}_2$.

a Demuestre que $\hat{\theta}_3$ es un estimador insesgado para θ.

b Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes, ¿cómo debe escogerse la constante a para minimizar la varianza de $\hat{\theta}_3$?

6 Considere la situación descrita en el Ejercicio 5. ¿Cómo debe elegirse la constante a para minimizar la varianza de $\hat{\theta}_3$, si $\hat{\theta}_1$ y $\hat{\theta}_2$ no son independientes pero son tales que $\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = c \neq 0$?

7 Suponga que Y_1, Y_2, Y_3 denotan una muestra aleatoria de una distribución exponencial con función de densidad

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta}\right) e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Considere los siguientes cinco estimadores de θ:

$$\hat{\theta}_1 = Y_1, \qquad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \qquad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \qquad \hat{\theta}_4 = \text{mín}(Y_1, Y_2, Y_3), \qquad \hat{\theta}_5 = \overline{Y}.$$

a ¿Cuáles de estos estimadores son insesgados?

b Entre los estimadores insesgados, ¿cuál tiene la varianza más pequeña?

8 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una población con función de densidad de probabilidad

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta + 1}\right) e^{-y/(\theta+1)}, & y > 0,\ \theta > -1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Sugiera un estadístico apropiado para usarlo como estimador insesgado para θ. [*Sugerencia:*considere$\overline{Y}$.]

9 Si Y tiene una distribución binomial con parámetros n y p, entonces $\hat{p}_1 = Y/n$ es un estimador insesgado de p. Otro estimador de p es $\hat{p}_2 = (Y + 1)/(n + 2)$.

a Deduzca el sesgo de $\hat{p}_2$.

b Deduzca $\text{MSE}(\hat{p}_1)$ y $\text{MSE}(\hat{p}_2)$.

c ¿Para qué valores de p es $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$?

10 Suponga que Y_1, Y_2, Y_3, Y_4 denotan una muestra aleatoria de tamaño 4 de una población con una distribución exponencial cuya densidad está dada por

$$f(y) \begin{cases} (1/\theta)e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Sea $X = \sqrt{Y_1 Y_2}$. Encuentre un múltiplo de X que sea un estimador insesgado para θ. [*Sugerencia:* use su conocimiento de la distribución gamma y el hecho de que $\Gamma(1/2) = \sqrt{\pi}$ para hallar $E(\sqrt{Y_1})$. Recuerde que las variables Y_i son independientes.]

b Sea $W = \sqrt{Y_1 Y_2 Y_3 Y_4}$. Encuentre un múltiplo de W que sea un estimador insesgado para θ^2. [Recuerde la sugerencia para el inciso a.]

11 Un investigador está interesado en la posibilidad de unir las aptitudes de televisión e Internet. Una muestra aleatoria de $n = 50$ usuarios de Internet dio que el tiempo medio semanal empleado en ver televisión era de 11.5 horas y que la desviación estándar era de 3.5 horas. Estime el tiempo medio poblacional que los usuarios de Internet pasan viendo televisión y fije un límite para el error de estimación.

12 Es frecuente que un aumento en el porcentaje de ahorros de los consumidores se encuentre ligado a la falta de confianza en la economía y se dice que es un indicador de una tendencia recesiva en aquélla. Un muestreo aleatorio de $n = 200$ cuentas de ahorros en una comunidad local mostró que el aumento medio en los valores de las cuentas de ahorros era de 7.2% en los últimos 12 meses, con desviación estándar de 5.6%. Estime el porcentaje medio de aumento en los valores de las cuentas de ahorros en los últimos 12 meses para depositantes de la comunidad. Establezca un límite para su error de estimación.

13 Los resultados de una encuesta de opinión pública publicados en la Internet[2] indicaron que 69% dequienes respondieron clasificaron el costo de la gasolina como una crisis o problema importante. El artículo indica que 1001 adultos, de 18 años o más, fueron entrevistados y que los resultados tienen un error de muestreo de 3%. ¿Cómo se calculó el 3% y cómo debería interpretarse? ¿Podemos concluir que una mayoría de los individuos del grupo de mayores de 18 años pensaron que el costo de la gasolina era una crisis o problema importante?

14 Se realizó un estudio para comparar el promedio de llamadas de emergencia a la policía en cada turno de 8 horas en dos distritos de una ciudad grande. Se seleccionaron aleatoriamente muestras de 100 turnos de 8 horas de los registros policíacos para cada una de las dos regiones y se registró el número de llamadas de emergencia para cada turno. Los estadísticos muestrales se proporcionan en la tabla siguiente.

	Región	
	1	2
Tamaño muestral	100	100
Media muestral	2.4	3.1
Varianza muestral	1.44	2.64

a Calcule la diferencia en el número medio de llamadas de emergencia a la policía por turno de 8 horas entre los dos distritos de la ciudad.

b Encuentre un límite para el error de estimación.

15 Podemos definir un límite de desviación estándar 2 en el error de estimación con cualquier estimador para el cual podamos hallar una estimación razonable del error estándar. Suponga que $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria de una distribución de Poisson con media λ. Sabemos que $V(Y_i) = \lambda$, y por tanto que $E(\overline{Y}) = \lambda$ y $V(\overline{Y}) = \lambda/n$. ¿Cómo emplearía usted $Y_1, Y_2, \ldots, Y_n$ para estimar λ? ¿Cómo estimaría el error estándar de su estimador?

16 Un auditor muestrea aleatoriamente 20 cuentas por cobrar de entre 500 de esas cuentas de la empresa de un cliente. El auditor hace una lista de la cantidad de cada cuenta y verifica si los documentos que sirven de base cumplen con los procedimientos estipulados. Los datos se registran en la tabla siguiente (las cantidades son en dólares, Y = sí y N = no).

Cuenta	Cantidad	Cumple	Cuenta	Cantidad	Cumple
1	278	Y	11	188	N
2	192	Y	12	212	N
3	310	Y	13	92	Y
4	94	N	14	56	Y
5	86	Y	15	142	Y
6	335	Y	16	37	Y
7	310	N	17	186	N
8	290	Y	18	221	Y
9	221	Y	19	219	N
10	168	Y	20	305	Y

Estime el total de cuentas por cobrar para las 500 cuentas de la empresa y fije un límite para el error de estimación. ¿Piensa usted que el *promedio* de cuentas por cobrar para la empresa es mayor a \$250 dólares? ¿Por qué?

17 Consulte el Ejercicio 16. De los datos obtenidos en las revisiones de cumplimiento, estime la proporción de las cuentas de la empresa que no cumplen con los procedimientos estipulados. Precise un límite para el error de estimación. ¿Piensa usted que la proporción de cuentas que cumplen con los procedimientos estipulados es mayor que 80%? ¿Por qué?

18 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución exponencial con media θ, entonces $E(Y_i) = \theta$ y $V(Y_i) = \theta^2$. Entonces, $E(\overline{Y}) = \theta$ y $V(\overline{Y}) = \theta^2/n$ o $\sigma_{\overline{Y}} = \theta/\sqrt{n}$. Sugiera un estimador insesgado para θ y dé una estimación para el error estándar del estimador sugerido.

19 Consulte el Ejercicio 18. Un ingeniero observa $n = 10$ mediciones independientes de la duración de un componente electrónico. El promedio de estas 10 mediciones es de 1020 horas. Si estas duraciones provienen de una distribución exponencial con media θ, estime θ y ponga un límite de error estándar de 2 en el error de estimación.

20 El número de personas que acuden a un banco de sangre hasta que se encuentra la primera de ellas con sangre tipo A es una variable aleatoria Y con distribución geométrica. Si p denota la probabilidad de que cualquier persona seleccionada aleatoriamente posea sangre tipo A, entonces $E(Y) = 1/p$ y $V(Y) = (1-p)/p^2$.

a Determine una función de Y que sea un estimador insesgado de $V(Y)$.

b Sugiera una forma de establecer un límite de error estándar de 2 para el error de estimación cuando Y se use para calcular $1/p$.