Trimestre Enero-Marzo 2008
Departamento de Cómputo Científico y Estadística
Guía de ejercicios. Estimación
Práctica N° 3

CONTENIDO
- Sesgo y error cuadrático medio de estimadores puntuales.
- Evaluación de la bondad de un estimador puntual.
- Intervalos de confianza.

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

a) Demuestre que $\widehat{\theta}_3$ es un estimador insesgado de θ.
b) Si $\widehat{\theta}_1$ y $\widehat{\theta}_2$ son independientes, ¿qué valor de la constante a debemos elegir para minimizar la varianza de $\widehat{\theta}_3$?

2. Suponga que Y_1, Y_2, Y_3 denotan una muestra aleatoria de una distribución exponencial cuya función de densidad es

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

Considere los siguientes cinco estimadores de θ :

$$\widehat{\theta}_1 = Y_1,\ \ \widehat{\theta}_2 = \tfrac{Y_1\,+\,Y_2}{2},\ \ \widehat{\theta}_3 = \tfrac{Y_1\,+\,2Y_2}{3},\ \ \widehat{\theta}_4 = mín(Y_1,\,Y_2,\,Y_3),\ \ \widehat{\theta}_5 = \overline{Y}$$

a) ¿Qué estimadores son insesgados?
b) Entre los estimadores insesgados, ¿cuál tiene menor varianza?

3. Suponga que $Y_1, Y_2, ..., Y_n$ constituyen una muestra aleatoria de una población cuya función de densidad de probabilidad es

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

a) Proponga un estadístico que se pueda usar como estimador insesgado de θ.

4. El número de descomposturas semanales de una minicomputadora es una variable aleatoria Y que tiene una distribución de $Poisson$ y media λ. Contamos con una muestra aleatoria $Y_1, Y_2, ..., Y_n$ de observaciones sobre la cantidad semanal de descomposturas.

a) Sugiera un estimador insesgado para λ.

b) El costo semanal de reparación de las descomposturas es $C = 3Y + Y^2$. Demuestre que $E(C) = 4\lambda + \lambda^2$.

c) Encuentre una función de $Y_1, Y_2, ..., Y_n$ que constituya un estimador insesgado de $E(C)$. Sugerencia: utilice lo que sabe de $\overline{Y}$ y $\left(\overline{Y}\right)^2$

5. La lectura de un voltímetro conectado a un circuito de prueba tiene una distribución uniforme en el intervalo $(\theta,\ \theta + 1)$, donde θ, cuyo valor se desconoce, es el voltaje

real del circuito. Suponga que $Y_1, Y_2, ..., Y_n$ denota una muestra aleatoria de tales lecturas.

a) Demuestre que $\overline{Y}$ es un estimador sesgado de θ y calcule el sesgo.

b) Encuentre una función de $\overline{Y}$ que sea un estimador insesgado de θ.

c) Encuentre $MS\overline{E}(Y)$ cuando se utiliza $\overline{Y}$ como estimador de θ.

6. Vimos que si Y tiene una distribución binomial con parámetros n y p, entonces Y/n es un estimador insesgado de p. Para estimar la varianza de Y por lo general usamos $n(Y/n)(1 - Y/n)$.

a) Demuestre que el estimador sugerido es un estimador sesgado de $Var(Y)$.

b) Modifique ligeramente $n(Y/n)(1 - Y/n)$ para formar un estimador insesgado de $Var(Y)$.

7. Sea $Y_1, Y_2, ..., Y_n$ una muestra aleatoria de tamaño n de una población cuya densidad está dada por

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

donde $\alpha > 0$ es un valor fijo conocido, pero el valor de θ se desconoce. Considere el estimador $\widehat{\theta} = máx(Y_1, Y_2, ..., Y_n)$.

a) Demuestre que $\widehat{\theta}$ es un estimador sesgado de θ.

b) Determine un múltiplo de $\widehat{\theta}$ que constituya un estimador insesgado de θ.

c) Deduzca $MSE(\widehat{\theta})$.

8. Suponga que $Y_1, Y_2, ..., Y_n$ es una muestra aleatoria de una distribución normal con parámetros μ y σ^2.

a) Demuestre que $S = \sqrt{S^2}$ es un estimador sesgado de σ. Sugerencia: recuerde la distribución de $(n-1)\,S^2/\sigma^2$ y el resultado del ejercicio 4.90.

b) Ajuste S para obtener un estimador insesgado de σ.

9. Suponga que $Y_1, Y_2, ..., Y_n$ denota una muestra aleatoria de tamaño n de una población con una distribución exponencial cuya densidad está dada por

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

Si $Y_{(1)} = mín(Y_1, Y_2, ..., Y_n)$ denota el menor estadístico de orden, demuestre que $\widehat{\theta} = nY_{(1)}$ es un estimador insesgado de θ, y determine $MSE(\widehat{\theta})$.

10. La Agencia para la protección del ambiente en conjunto con la Universidad de Florida recientemente realizó un amplio estudio respecto al posible efecto de los oligoelementos presentes en el agua potable en la formación de cálculos renales. En la siguiente tabla se presentan los datos, los cuales se obtuvieron de individuos con problemas recurrentes de cálculos renales que viven en los estados de las dos Carolinas y las Montañas Rocallosas, respecto a la edad, la concentración de calcio en el agua potable (medida en partes por millón) y el hábito de fumar.

	Carolinas	Montañas Rocallosas
Tamaño de la muestra	467	191
Edad promedio	45.1	46.4
Desviación estándar de la edad	10.2	9.8
Concentración promedio de calcio	11.3	40.1
Desviación estándar del calcio	16.6	28.4
Proporción de fumadores en el momento del estudio	0.78	0.61

a) Estime la concentración promedio de calcio en el agua potable para los pacientes con cálculos que viven en las Carolinas. Establezca un límite para el error de estimación.

b) Estime la diferencia en la media de edades de los pacientes con cálculos renales en las Carolinas y en las Montañas Rocallosas. Establezca un límite para el error de estimación.

c) Estime y precise un límite de dos desviaciones estándar para la diferencia en las proporciones del número de pacientes con cálculos renales que viven en las Carolinas y en las Rocallosas que fumaban en el momento en que se realizó el estudio.

11. Los resultados de un sondeo de opinión pública, publicados en una revista de noticias (*Time*, 5 de abril de 1993) indica que el 54% de los encuestados entre las edades de 18 y 26 años consideraban que la religión constituye una parte “muy importante” de sus vidas. El artículo afirma que se entrevistó a 1013 individuos y que los resultados tienen un error de muestreo de 3%. ¿Cómo se calculó el 3%, y cómo debe interpretarse? ¿Podemos concluir que la mayoría de los individuos en este grupo de edades considera que la religión es una parte muy importante de sus vidas?

12. Se realizó un estudio para comparar el promedio de llamadas telefónicas de emergencia que entran a las oficinas de la policía en cada turno de 8 horas en dos distritos de una gran ciudad. Se seleccionaron aleatoriamente 100 muestras de turnos de 8 horas de los registros de la policía de cada una de las dos regiones, y se registró el número de llamadas de emergencia que se recibieron en cada turno. La siguiente tabla contiene las estadísticas de la muestra.

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

a) Estime la diferencia en la media del número de llamadas de emergencia que entraron en cada turno de 8 horas en los dos distritos de la ciudad. Precise un límite para el error de estimación.

13. Durante más de 15 años Louis Harris y Asociados han llevado a cabo encuestas sobre los hábitos relacionados con la salud de los estadounidenses. Las más recientes revelaron hallazgos que el doctor Todd Davis, vicepresidente de la asociación de médicos de Estados Unidos, calificó como “muy preocupantes”. La siguiente tabla contiene los resultados de las encuestas realizadas en 1983 y 1992, la encuesta de 1992 incluía $n = 1251$ individuos. Aunque el tamaño de la encuesta de 1983 no se indica en el artículo, suponga que incluía a $n = 1250$ individuos.

Estado de salud	Encuesta de 1983	Encuesta de 1992
Consumían la cantidad prescrita de alimentos ricos en fibra	0.59	0.53
Evitaban las grasas	0.55	0.51
Evitaban el exceso de sal	0.53	0.46
Utilizaban cinturón de seguridad	0.19	0.70
Contaban con detectores de humo	0.67	0.90

a) Indique una estimación puntual de la diferencia en las proporciones de estadounidenses que consumían la cantidad prescrita de alimentos ricos en fibra en 1983 y 1992. Precise un límite para el error de estimación.

b) Con base en los resultados del inciso a), ¿considera usted que disminuyó de manera evidente el número de estadounidenses que consumía la cantidad prescrita de alimentos ricos en fibras entre 1983 y 1992? Explique.

14. Repase el ejercicio 13. Los resultados de la encuesta de Harris aplicada en 1992 contienen buenas noticias respecto al número de individuos que acostumbran usar el cinturón de seguridad. Indique la estimación puntual y un límite para el error de estimación de la proporción de personas que utilizaba casi siempre el cinturón de seguridad en 1992. ¿Es probable que el valor de su estimación difiera del valor real en 10%? ¿Por qué?

15. Podemos definir un límite de dos desviaciones estándar para el error de estimación con cualquier estimador para el cual podamos determinar una estimación razonable del error estándar. Suponga que $Y_1, Y_2, ..., Y_n$ representa una muestra aleatoria de una distribución de Poisson con media λ. Sabemos que $Var(Y_i) = \lambda$ y, por consiguiente, que $E(\overline{Y}) = \lambda$ y $Var(\overline{Y}) = \lambda/n$. ¿Cómo emplearía $Y_1, Y_2, ..., Y_n$ para estimar λ? ¿Cómo estimaría el error estándar de su estimador?

16. La cantidad de puntos de nucleación de gérmenes de cristales por unidad de volumen en el aluminio policristalino tiene una distribución de Poisson con media λ. Cincuenta muestras de volumen unitario en las pruebas de recocido de acuerdo con el método A revelaron un promedio de 20 puntos por unidad de volumen. 50 muestras de volumen unitario elegidas de manera independiente en las pruebas de recocido con el método B revelaron un promedio de 23 puntos por unidad de volumen.

a) Estime la media λ_A del número de puntos de nucleación con el método A, y precise un límite de error de dos desviaciones estándar para el error de estimación.

b) Estime la diferencia en la media del número de puntos de nucleación $\lambda_A - \lambda_B$ con los métodos A y B. Establezca un límite de dos desviaciones estándar para el error de estimación. ¿Diría usted que el método B produce un promedio mayor de puntos de nucleación? ¿Por qué?

17. Suponga que la variable aleatoria Y es una observación de una distribución normal con media desconocida μ y varianza 1.

a) Encuentre un intervalo de confianza de 95% para μ.
b) Encuentre un límite de confianza superior de 95% para μ.
c) Encuentre un límite de confianza inferior de 95% para μ.

18. Sea $Y_1, Y_2, ..., Y_n$ una muestra aleatoria de tamaño n de una población con distribución uniforme en el intervalo $(0, \theta)$. Si $Y_{(n)} = m\acute{a}x(Y_1, Y_2, ..., Y_n)$ y $U = (1/\theta)\, Y_{(n)}$:

a) Demuestre que U tiene la siguiente función de distribución:

$$F_U(u) = \begin{cases} 0, & u < 0 \\ u^n, & 0 \leq u \leq 1 \\ 1, & u > 1 \end{cases}$$

b) Como la distribución de U no depende de θ, U es una cantidad pivote. Encuentre un límite de confianza inferior de 95% para θ.

19. Las longitudes de los caparazones de 10 langostas analizados en un estudio de la infestación de la langosta *Thenus orientalis* por dos tipos de percebes, *Octolasmis tridens* y *O. lowei*, aparecen en la siguiente tabla. Encuentre un intervalo de confianza de 95% para la longitud media del caparazón de las langostas *Thenus orientalis* capturadas en las aguas de los alrededores de Singapur [W.B. Jeffries, H. K. Voris y C. M. Yang: "Diversity and Distribution of the Pedunculate Barnacle *Octolasmis* Gray, 1825 Epizoic on the Scyllarid Lobsters, *Thenus orientalis* (Lund, 1793)" *Crustaceana*, 46, núm. 3, 1984].

Número de sector de la langosta	A061	A062	A066	A070	A067
Longitud del caparazón (mm)	78	66	65	63	60

Número de sector de la langosta	A069	A064	A068	A065	A063
Longitud del caparazón (mm)	60	58	56	52	50

20. Los resultados de la prueba de conocimientos generales (SAT por sus siglas en inglés), que han disminuido lentamente desde el inicio de la prueba, ahora han comenzado a elevarse. Al principio se contempló un promedio de 500 puntos. En 1991 el promedio fue de aproximadamente 422 puntos para el examen de habilidad verbal y 474 puntos para el de matemáticas. Una muestra aleatoria de los resultados de 20 estudiantes de último año de una preparatoria urbana grande dio como resultado las medias y las desviaciones estándar que aparecen en la tabla siguiente.

	Capacidad verbal	Matemáticas
Media de la muestra	419	455
Desviación estándar de la muestra	57	69

a) Encuentre un intervalo de confianza de 90% para la media de los resultados del examen verbal de los estudiantes de último grado de la escuela preparatoria.

b) ¿El intervalo que encontró en el inciso a) incluye el valor de 422, la media real de los resultados del examen de capacidad verbal aplicado en 1991? ¿Qué concluye usted?

c) Construya un intervalo de confianza del 90% para la media real de los resultados del examen de matemáticas que presentaron los estudiantes del último grado de la preparatoria. ¿Incluye este intervalo el valor de 474, la media real de los resultados de matemáticas de 1991? ¿Qué puede concluir de esto?

21. El síndrome del comportamiento anterior crónico es un mal que se manifiesta por un dolor en el muslo debido al ejercicio. Inflamación y daño al nervio y a la función muscular se suman al dolor, que disminuye con reposo. Susan Beckham y colaboradores (1993) realizaron un experimento que incluía a 10 corredores y a 10 ciclistas saludables para determinar si las mediciones de la presión sobre el comportamiento del musculo anterior difieren en los corredores y en los ciclistas. En la siguiente tabla se resumen los datos (la presión ejercida sobre el comportamiento se mide en milímetros de mercurio)

	Corredores		Ciclistas	
Estado	Media	s	Media	s
Reposo	14.5	3.92	11.1	3.98
80% de consumo máximo de O_2	12.2	3.49	11.5	4.95

a) Construya un intervalo de confianza de 95% para la diferencia en las presiones medias sobre el comportamiento entre los corredores y los ciclistas que están en reposo.

b) Construya un intervalo de confianza de 90% para la diferencia en las presiones medias sobre el compartimiento entre los corredores y los ciclistas que hacen ejercicio con un 80% de consumo máximo de oxígeno.

22. Se administran dos nuevos medicamentos a pacientes que padecen hipertensión. El primero redujo la presión sanguínea de 16 pacientes un promedio de 11 puntos, con una desviación estándar de 6 puntos. El segundo redujo la presión sanguínea de otros 20 pacientes un promedio de 12 puntos, con una desviación estándar de 8 puntos. Encuentre un intervalo de confianza de 95% para la diferencia en las reducciones medias de la presión sanguínea, suponiendo que las mediciones tienen una distribución normal con varianzas iguales.

23. La Agencia para la Protección del Ambiente de Estados Unidos reunió datos de mediciones de LC50 (concentraciones que matan 50% de los animales de experimentación) de ciertos productos químicos que posiblemente se encuentren en ríos y lagos de agua dulce. Las mediciones de LC50 (en partes de millón) para el DDT en 12 experimentos con cierta especie de peces fueron las siguientes:

16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9

Estime la media real de LC50 para el DDT, con un coeficiente de confianza de 0.90. Suponga que las mediciones de LC50 tienen una distribución aproximadamente normal.

24. Consulte con el ejercicio 23. Otro insecticida común, el Diazinon, dio las siguientes mediciones de CL50 en tres experimentos: 7.8, 1.6 y 1.3.

a) Estime la media de LC50 para el Diazinon con un intervalo de confianza de 90%.
b) Estime la diferencia entre la media de LC50 para el DDT y la media de LC50 para el Diazinon con un intervalo de confianza de 90% ¿Qué suposiciones es necesario hacer para que sea válido el método que usted utilizó?

25. Una fábrica trabaja con dos máquinas, A y B. El costo semanal X de reparación de las máquinas tipo A tiene una distribución normal con media μ_1 y varianza σ^2 . El costo semanal Y de reparación de las máquinas tipo B también tiene una distribución normal con media μ_2 y varianza $3\sigma^2$. El costo semanal esperado para la fábrica es, por consiguiente, de $2\mu_1 + \mu_2$. Si tenemos una muestra aleatoria $X_1, \ldots , X_n$ de costos para las máquinas tipo A y una muestra aleatoria independiente $Y_1,\ldots ,Y_m$, de costos para las máquinas tipo B, describa cómo construiría un intervalo de confianza de 95% para $2\mu_1 + \mu_2$ si:

a) se conoce el valor de σ^2.
b) se desconoce el valor de σ^2.

26. Recientemente, la Agencia para la Protección del Ambiente de Estados Unidos estableció que el nivel máximo de ruido que pueden hacer los camiones de carga es de 83 decibeles (dB). La forma en que se aplique este límite afectará considerablemente al público y a la industria camionera. Una forma de aplicarlo es exigir que todos los camiones respeten el límite de ruido. Otro método (aunque menos eficaz) consiste en exigir que el nivel promedio de ruido de una flotilla de camiones no sobrepase el límite. En el caso de que se adopte esta última regla, la variación del nivel de ruido de un camión a otro se torna importante, ya que un valor grande de σ^2 implicaría que muchos camiones rebasen el límite, aún si el nivel medio de ruido que hace una flotilla es de 83 dB. Una muestra aleatoria de seis camiones de carga produjo los siguientes niveles de ruido (en decibeles):

85.4 86.8 86.1 85.3 84.8 86.0

Construya con estos datos un intervalo de confianza de 90% para σ^2, la varianza de las lecturas de emisión de ruidos que hacen los camiones. Interprete sus resultados.

27. Suponga que S^2 es la varianza muestral basada en una muestra de tamaño n de una población normal cuya media y varianza se desconocen.

a) Obtenga un límite de confianza superior de $100\,(1-\alpha)\,\%$ para σ^2.
b) Obtenga un límite de confianza inferior de $100\,(1-\alpha)\,\%$ para σ^2.

28. En el ejercicio anterior usted dedujo límites de confianza superior e inferior, cada uno con un coeficiente de confianza de $1-\alpha$ para σ^2.

a) ¿Cómo generaría un límite de confianza superior de $100\,(1-\alpha)\,\%$ para σ?
b) ¿Cómo generaría un límite de confianza inferior de $100\,(1-\alpha)\,\%$ para σ?

29. Las edades de una muestra aleatoria de cinco profesores universitarios son de 39, 54, 61, 72 y 59. Con esta información encuentre un intervalo de confianza de 99% para la desviación estándar de la población de las edades de los profesores de la universidad, suponiendo que éstas tienen una distribución normal estándar.