Probabilidad y Estadística (93.24)
Trabajo Práctico N$^{\underline{o}}$ 8

Estimación de parámetros.

1. El tiempo de duración de una pieza de un equipo puede considerarse una variable aleatoria normal con desvío estándar de 4 horas y una media μ que se desea estimar. Una muestra aleatoria de 100 piezas que fueron probadas produjo una media muestral de 501.2 horas de duración. Obtener un intervalo de confianza para la media μ con un nivel de confianza de: a) 0.95 (95 %), b) 0.99 (99 %).

2. En una fábrica de materiales eléctricos se desea estimar el peso promedio del último lote de rollos de alambre de cobre salido de producción. Para ello se eligió al azar una muestra de 20 rollos que arrojó un promedio de 38 Kg por rollo. Se conoce además, de registros históricos, que el desvío estándar del peso de un rollo es de 4.2 Kg.

   a) Estimar el peso medio de los rollos con un intervalo de confinza del 95 %.

   b) ¿Cuántos rollos más habría que pesar para poder obtener una estimación del peso medio de un rollo con un error de muestreo de 1 Kg?

3. Una máquina llenadora de latas de café dosifica cantidades variables con distribución normal de desvío estándar de 15 gramos. A intervalos regulares se toman muestras de 10 envases con el fin de estimar la dosificación media. Una de estas muestras arrojó una media de 246 gramos.

   a) Obtener un intervalo de confianza del 90 % para la dosificación media.

   b) ¿Cuántos envases más habría que pesar para poder obtener una estimación cuyo error de muestreo fuera 5 gramos?

4. Se conectan 25 lámparas de luz infrarroja en un invernadero, de tal manera que si falla una lámpara, otra se enciende inmediatamente (se enciende solamente una lámpara a la vez). Las lámparas funcionan en forma independiente y el tiempo de duración puede suponerse una variable aleatoria normal con una vida media de 50 horas y una desviación estándar de 4 horas. Si T es el tiempo total de operación de las 25 lámparas de luz infrarroja,

   a) hallar la probabilidad de que dicho tiempo T exceda 1300 horas. Explicar el procedimiento utilizado detallando la variable aleatoria con la que se calcula la probabilidad solicitada.

   b) La variable aleatoria $T/25$ puede considerarse como una estimación puntual de la media de la duración de una lámpara en base a una muestra de tamaño 25. Obtener un intervalo de confianza del 95 % para la vida media de una lámpara si el tiempo total de operación fue 1251 horas.

5. Sea X una variable aleatoria con media μ y varianza σ^2. Considere dos muestras aleatorias independientes de tamaños n_1 y n_2, con medias muestrales $\overline{X}_1$ y $\overline{X}_2$. Se define el estadístico

$$Y = a\,\overline{X}_1 + b\,\overline{X}_2$$

donde $0 < a < 1$.

   a) Determine b en función de a para que Y sea un estimador insesgado de μ.
   b) Utilizando la relación entre b y a hallada en el ítem anterior, encuentre el valor de a que minimiza la varianza de Y si $n_2 = 2\,n_1$.
   c) Suponga que $n_2 = n_1 = n$ y $b = 1 - a$. Demuestre que $\mathrm{Var}(Y) < \mathrm{Var}(\overline{X}_1) = \mathrm{Var}(\overline{X}_2)$ si $a \in (0,1)$ y que el mínimo de $\mathrm{Var}(Y)$ es $\sigma^2/2\,n$ para $a = \frac{1}{2}$.

6. Se sabe que determinaciones hechas sobre la densidad de cierto producto químico se distribuyen normalmente alrededor de la media poblacional desconocida. Esa densidad puede considerarse una variable aleatoria con distribución normal con un desvío estándar de 0.005 g/cm^3. Se desea estimar la densidad media con un intervalo de confianza del 95 % y con un error menor que 0.002 g/cm^3. ¿Cuál deberá ser el tamaño mínimo de la muestra?

7. La media muestral y la dispersión muestral de la carga máxima soportada por un cable en una muestra de tamaño 60 son 11.09 y 0.73 Tn respectivamente. Hallar los límites de confianza del 95 % para la carga máxima media.

8. En la siguiente tabla se presentan los datos del contenido de silicio en una muestra de 150 coladas de hierro:

Contenido de sílice	**Cantidad de coladas**
0.333 - 0.433	4
0.433 - 0.533	12
0.533 - 0.633	19
0.633 - 0.733	28
0.733 - 0.833	48
0.833 - 0.933	25
0.933 - 1.033	14

Estimar con una confianza del 95 % el contenido medio de sílice por colada.

9. Los contenidos de 7 recipientes similares para ácido sulfúrico son: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 y 9.6 litros. Obtener intervalos de confianza del 95 % para la media y la dispersión del contenido de los recipientes de esa clase asumiendo que el contenido de ácido en los recipientes es una variable aleatoria normal.

10. En una industria textil hay un lote de tambores de 100 litros de capacidad que contienen un suavizante textil, que se han usado parcialmente, por lo que se desea estimar el contenido medio de los mismos. A tal efecto se tomó una muestra aleatoria de 15 tambores, se midieron sus

contenidos y se obtuvo una media muestral de 63 litros con un desvío estándar muestral de 12.5 litros. Determinar los límites de confianza del 90 % para el contenido medio de los tambores de la población. Suponga que el contenido de suavizante en los recipientes tiene distribución normal.

11. Se hicieron pruebas de inmersión en solución corrosiva de una aleación determinándose el porcentaje de pérdida en resistencia a la tracción (que se supone una variable aleatoria con distribución normal). Los resultados obtenidos fueron: 6.4, 4.6, 4.6, 6.4, 3.2, 5.2, 6.5, 4.9, 4.3, 5.6, 3.7 y 4.6. Encontrar un intervalo de confianza del 95 % para el porcentaje esperado de pérdida en resistencia a la tracción.

12. De lo producido en cierto proceso se tomó una muestra aleatoria de 6 artículos y se midió una longitud característica (que se supone una variable aleatoria con distribución normal), obteniéndose los siguientes valores (en cm): 2.9, 2.92, 2.9, 2.84, 2.79 y 2.8. Determinar un intervalo de confianza del 95 % para la longitud característica media.

13. Las ventas de una revista semanal han sido las siguientes (en miles de ejemplares) en las últimas 4 semanas: 15.4, 18.5, 16.3, y 19.2. Calcular los límites de confianza del 95 % para el promedio y para la varianza de las ventas semanales (se supone que el volumen semanal de ventas es una variable aleatoria con distribución normal).

14. Una máquina produce varillas de metal usadas en el sistema de suspensión de un auto. Se toma una muestra aleatoria de 15 varillas y se miden los diámetros. Los datos obtenidos (en mm) se detallan más abajo. Asumiendo que el diámetro de las varillas está distribuido normalmente obtener un intervalo de confianza del 95 % para el diámetro medio de la varilla.

Diam. (mm)	8.24	8.23	8.20	8.21	8.20	8.28	8.23	8.26
	8.24	8.24	8.25	8.19	8.25	8.26	8.23	

15. De un proceso productivo de una pieza seriada se tomó una muestra de 300 unidades en la que se encontraron 18 defectuosas.

    a) Calcular los límites de confianza del 90 % para el porcentaje defectuoso del proceso.

    b) Calcular el tamaño de muestra adicional para tener un intervalo del mismo nivel de confianza pero de semiamplitud 0.01 (o sea del 1 % de semiamplitud).

    c) Con la muestra dada de 300 unidades calcular el porcentaje defectuoso máximo del proceso con 90 % de confianza (o sea un porcentaje tal que la probabilidad de que el verdadero porcentaje defectuoso lo exceda sea 0.1).

16. Una muestra de 100 votantes elegidos al azar indicó que 55 % de ellos estaba a favor de un candidato. Obtener un intervalo de confianza del 95 % para la proporción de votantes que en el total de la población votan por este candidato.

17. El *rating* de un programa de televisión se mide como el porcentaje de hogares que está viendo el programa en un momento dado. Una compañía medidora de rating cuenta con un panel de 600 hogares colaboradores, en los cuales ha instalado un *people meter* (dispositivo que registra cada minuto si el televisor está encendido y en qué canal, y envía telefónicamente la información a la base de datos). Se ha registrado el rating del programa *Bailando por bailar* en 25 puntos. Es decir que el 25 % de los hogares del panel vió todo el programa ese día.

    a) Calcular un intervalo de confianza del 90 % para el rating del programa.

    b) Sabiendo que cada people meter cuesta \$ C, calcular la inversión adicional en dispositivos necesaria para medir el rating con un error de muestreo de ± 1 %.

18. En un diario de gran circulación de esta ciudad se publicó el resultado de una encuesta en la que se indicaban, para una muestra de tamaño n, varias *proporciones* de preferencia por varios equipos de fútbol. En ese reporte figuraba una ficha técnica en la que se indicaba lo siguiente:

    - Tamaño de la muestra $n = 9300$
    - Nivel de confianza 95 %
    - Error máximo: 1 % (semiamplitud máxima de un intervalo de confianza).

    Explicite claramente un planteo en el cuál dadas dos de las tres cantidades indicadas se obtenga la tercera. Verificar, por ejemplo, que para ese nivel de confianza y tamaño muestral el error máximo es aproximadamente 0.01 (1 %).
    En esa misma encuesta se indicaba que *los hinchas de Boca no son la mitad mas uno* dado que el 41 % (0.403) de la muestra correspondió a hinchas de Boca. Explique el significado de esta afirmación a partir del intervalo de confianza hallado.

19. Considere una cierta población de tamaño finito N en la que se encuentra definido un atributo A en sus integrantes. Un parámetro poblacional de interés es la proporción p de integrantes de la población que tienen el atributo. Un estimador insesgado de p es $\hat{p} = \frac{X}{n}$, donde X es el número de miembros de una muestra de tamaño n, extraída de esta población de tamaño N, que tienen el atributo bajo análisis. Si la muestra se extrae sin reposición y vale la aproximación normal de la distribución hipergeométrica, entonces se puede demostrar que la semiamplitud de un intervalo de confianza del $100\,(1-\alpha)$ % para p, viene dada por

    $$E = z_{1-\frac{\alpha}{2}}\,\sqrt{\hat{p}\,(1-\hat{p})}\,\sqrt{\frac{N-n}{N\,n}}.$$

    a) Verificar que si $N \gg n$ entonces se obtiene el resultado para poblaciones infinitas. Analizar el caso en que $N = n$ (la muestra en este caso es una sola posible).

    b) Sea $B = \frac{E}{z_{1-\frac{\alpha}{2}}\,\sqrt{\hat{p}\,(1-\hat{p})}}$. Probar que el tamaño de la muestra necesario para estimar p, con un nivel de confianza $100\,(1-\alpha)$ %, viene dado por

    $$n = \frac{1}{B^2 + \frac{1}{N}}.$$

    En este caso el valor de $\hat{p}$ es el que por ejemplo corresponda a una estimación *a priori* de p (por ejemplo el resultado de una muestra piloto).

c) Analizar n como función de N si por ejemplo $E = 0.01$ ($1\,\%$ es la semiamplitud del intervalo de confianza para p), $z = 2$ (o $(1-\alpha) \approx 0.955$) y $\hat{p} = 0.5$. Usar para N valores del la forma 10^k con k tomando los valores $2, 3, \ldots, 7$. Extraiga alguna conclusión.

Apéndice: Una simulación en Octave.
Optativo y recomendable para fijar conceptos.

Se sugiere ejecutar los códigos de *Octave* donde se repite N veces la operación de extraer una muestra de tamaño n y en esa muestra se evalúa un estimador. De esta manera se tienen N realizaciones del estadístico y se puede realizar una estadística de ese conjunto de valores.

a) La secuencia que sigue a continuación se refiere al problema de la estimación de una proporción poblacional a partir de una muestra de tamaño n:

```
p = rand; n = 100; N = 1000; X = (rand(n,N)<p);
M = mean(X);
hist(M)
[ mean(M) p std(M) sqrt(p*(1-p)/n)]
```

Breve explicación: Se genera el valor de `p` con un número aleatorio, luego se obtiene la matriz `X` de `n` filas y `N` columnas con 1 o 0 dependiendo que la condición entre paréntesis sea verdadera o falsa, respectivamente. Entonces la proporción de 1 (en el largo plazo o para n grande) es `p`. El vector `M` almacena `N` valores de la media (la proporción muestral $\hat{p}$) de una muestra de tamaño `n` de cada columna de `X`. El comando `hist` genera el histograma de los N valores del vector `M`. Luego se imprimen cuatro números: el promedio de los `n` valores de `M`, el valor de `p`, el desvío estándar del vector `M`, y el desvío estándar del estimador $\hat{p}$). Una ejecución de este código devolvió estos cuatro números 0.15451, 0.15377, 0.036667 y 0.036073.

b) El código que sigue se refiere al problema de estimar la media de una variable aleatoria normal a partir de una muestra de tamaño n:

```
mu = 100*rand;sigma = 0.1*mu;
n = 100; N = 1000; X=mu+sigma*randn(n,N);
M = mean(X);
hist(M)
mean(M)
[ mean(M) mu std(M) sigma/sqrt(n)]
```

Breve explicación: Se generan los valores de μ y σ, el primero como 100 veces un número aleatorio, y σ como el 10 % de μ. Luego se obtiene la matriz `X` de `n` filas y `N` columnas con números aleatorios provenientes de una variable aleatoria normal con la media y el desvío generados. El vector `M` almacena `N` valores de la media (la media muestral $\hat{\mu} = \overline{x}$) de una muestra de tamaño `n` de cada columna de `X`. El comando `hist` genera el histograma de los N valores del vector `M`. Luego se imprimen cuatro números: el promedio de los `n` valores de `M`, el valor de μ, el desvío estándar del vector `M`, y el desvío estándar del estimador $\hat{\mu}$. Una ejecución de este código devolvió estos cuatro números 71.611, 71.579, 0.71086 y 0.71579.

c) A continuación se generan N intervalos de confianza para la estimación de la media μ de una variable aleatoria normal a partir de una muestra de tamaño n, y se mide la frecuencia de intervalos que tienen la propiedad de contener el valor de μ:

```
mu = 100*rand; sigma = 0.1*mu;
```

```
n = 100; N = 1000; X=mu + sigma*randn(n,N);
M = mean(X);
A = M - 1.6449*sigma/sqrt(n); B = M + 1.6449*sigma/sqrt(n);
p = mean((A < mu).*(mu < B))
```

Breve explicación: Se generan los valores de μ y σ, el primero como 100 veces un número aleatorio, y σ como el 10 % de μ. Luego se obtiene la matriz `X` de `n` filas y `N` columnas con números aleatorios provenientes de una variable aleatoria normal con la media y el desvío generados. El vector `M` almacena `N` valores de la media (la media muestral $\hat{\mu} = \overline{x}$) de una muestra de tamaño `n` de cada columna de `X`. Los vectores `A` y `B` contienen los `N` extremos izquierdo y derecho de intervalos de confianza del 90 % con centro en cada uno de los valores generados de la media muestral obtenida en base a muestras de tamaño `n`. El valor de `p` es la proporción de esos intervalos que contienen el valor de μ. Una ejecución de este código devolvió `p = 0.90900`.