## Tema 1: Introducción a la Inferencia Estadística. Estadísticos muestrales

1. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$. Dar el espacio muestral y calcular la función masa de probabilidad de $(X_1, \ldots, X_n)$ en cada uno de los siguientes casos:

   *a)* $X \to \{B(k_0, p);\quad p \in (0,1)\}$ Binomial

   *b)* $X \to \{\mathcal{P}(\lambda);\quad \lambda \in \mathbb{R}^+\}$ Poisson

   *c)* $X \to \{BN(k_0, p);\quad p \in (0,1)\}$ Binomial Negativa

   *d)* $X \to \{G(p);\quad p \in (0,1)\}$ Geométrica

   *e)* $X \to \{P_N;\quad N \in \mathbb{N}\},\quad P_N(X = x) = \dfrac{1}{N},\quad x = 1, \ldots, N.$

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$. Dar el espacio muestral y calcular la función de densidad de $(X_1, \ldots, X_n)$ en cada uno de los siguientes casos:

   *a)* $X \to \{U(a,b);\quad a, b \in \mathbb{R},\ a < b\}$ Uniforme

   *b)* $X \to \{\mathcal{N}(\mu, \sigma^2);\ \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$ Normal

   *c)* $X \to \{\Gamma(p, a);\quad p, a \in \mathbb{R}^+\}$ Gamma

   *d)* $X \to \{\beta(p, q);\quad p, q \in \mathbb{R}^+\}$ Beta

   *e)* $X \to \{P_\theta;\quad \theta \in \mathbb{R}^+\},\quad f_\theta(x) = \dfrac{1}{2\sqrt{x\theta}},\quad 0 < x < \theta.$

3. Se miden los tiempos de sedimentación de una muestra de partículas flotando en un líquido. Los tiempos observados son: 11.5, 1.8, 7.3, 12.1 1.8, 21.3, 7.3, 15.2, 7.3, 12.1, 15.2, 7.3, 12.1, 1.8, 10.5, 15.2, 21.3, 10.5, 15.2, 11.5.

   - Construir la función de distribución muestral asociada a a dichas observaciones.
   - Hallar los valores de los tres primeros momentos muestrales respecto al origen y respecto a la media.
   - Determinar los valores de los cuartiles muestrales.

4. Se dispone de una muestra aleatoria simple de tamaño 40 de una distribución exponencial de media 3, ¿cuál es la probabilidad de que los valores de la función de distribución muestral y la teórica, en $x = 1$, difieran menos de 0.01? Aproximadamente, ¿cuál debe ser el tamaño muestral para que dicha probabilidad sea como mínimo 0.98?

5. Se dispone de una muestra aleatoria simple de tamaño 50 de una distribución de Poisson de media 2, ¿cuál es la probabilidad de que los valores de la función de distribución muestral y la teórica, en $x = 2$, difieran menos de 0.02? Aproximadamente, ¿qué tamaño muestral hay que tomar para que dicha probabilidad sea como mínimo 0.99?

6. Sea $X \to B(1, p)$ y $(X_1, X_2, X_3)$ una muestra aleatoria simple de $X$. Calcular la función masa de probabilidad de los estadísticos $\overline{X},\ S^2,\ \min X_i$ y $\max X_i$.

7. Obtener la función masa de probabilidad o función de densidad de $\overline{X}$ en el muestreo de una variable de Bernoulli, de una Poisson y de una exponencial.

8. Calcular las funciones de densidad de los estadísticos $\max X_i$ y $\min X_i$ en el muestreo de una variable $X$ con función de densidad:
$$f_\theta(x) = e^{\theta - x}, \quad x > \theta.$$

9. El número de pacientes que visitan diariamente una determinada consulta médica es una variable aleatoria con varianza de 16 personas. Se supone que el número de visitas de cada día es independiente de cualquier otro. Si se observa el número de visitas diarias durante 64 días, calcular aproximadamente la probabilidad de que la media muestral no difiera en más de una persona del valor medio verdadero de visitas diarias.

10. Una máquina de refrescos está arreglada para que la cantidad de bebida que sirve sea una variable aleatoria con media 200 ml. y desviación típica 15 ml. Calcular de forma aproximada la probabilidad de que la cantidad media servida en una muestra aleatoria de tamaño 36 sea al menos 204 ml.

## Tema 2: Distribuciones en el muestreo de poblaciones normales

1. Se toma una muestra aleatoria simple de tamaño 5 de una variable aleatoria con distribución $\mathcal{N}(2.5, 36)$. Calcular:

   a) Probabilidad de que la cuasivarianza muestral esté comprendida entre 1.863 y 2.674.

   b) Probabilidad de que la media muestral esté comprendida entre 1.3 y 3.5, supuesto que la cuasivarianza muestral está entre 30 y 40.

2. La longitud craneal en una determinada población humana es una variable aleatoria que sigue una distribución normal con media 185.6 mm. y desviación típica 12.78 mm. ¿Cuál es la probabilidad de que una muestra aleatoria simple de tamaño 20 de esa población tenga media mayor que 190 mm.?

3. ¿De que tamaño mínimo habría que seleccionar una muestra de una variable con distribución normal $\mathcal{N}(\mu, 4)$ para poder afirmar, con probabilidad mayor que 0.9, que la media muestral diferirá de la poblacional menos de 0.1?

4. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable con distribución normal. Calcular la probabilidad de que la cuasivarianza muestral sea menor que un 50 % de la varianza poblacional para $n = 16$ y para $n = 1000$.

5. Sean $S_1^2$ y $S_2^2$ las cuasivarianzas muestrales de dos muestras independientes de tamaños $n_1 = 5$ y $n_2 = 4$ de dos poblaciones normales con la misma varianza. Calcular la probabilidad de que $S_1^2/S_2^2$ sea menor que 5.34 o mayor que 9.12.

6. Se consideran dos poblaciones de bombillas cuyas longitudes de vida siguen una ley normal con la misma media y desviaciones típicas 425 y 375 horas, respectivamente. Con objeto de realizar un estudio comparativo de ambas poblaciones, se considera una muestra aleatoria simple de 10 bombillas en la primera población y una de tamaño 6 en la segunda. ¿Cuál es la probabilidad de la diferencia entre las medias muestrales del primer y segundo grupo sea menor que la observada en dos realizaciones muestrales que dieron 1325 horas y 1215 horas, respectivamente?

7. Sean $X_1, \ldots, X_n, X_{n+1}$ variables aleatorias independientes e idénticamente distribuidas según una $\mathcal{N}(\mu, \sigma^2)$, y sean $\overline{X}$ y $S^2$ la media y cuasivarianza muestral de $(X_1, \ldots, X_n)$. Calcular la distribución de
$$\frac{X_{n+1} - \overline{X}}{S}\sqrt{\frac{n}{n+1}}.$$

8. Sean $(X_1, \ldots, X_n)$, $(Y_1, \ldots, Y_m)$ muestras aleatorias simples independientes de poblaciones $\mathcal{N}(\mu_1, \sigma^2)$ y $\mathcal{N}(\mu_2, \sigma^2)$, respectivamente. Sean $\alpha$, $\beta \in \mathbb{R}$, y $\overline{X}$, $\overline{Y}$, $S_1^2$, $S_2^2$ las medias y cuasivarianzas de las dos muestras. Calcular la distribución de
$$\frac{\alpha(\overline{X} - \mu_1) + \beta(\overline{Y} - \mu_2)}{\sqrt{\dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}\sqrt{\dfrac{\alpha^2}{n} + \dfrac{\beta^2}{m}}}.$$

## Tema 3: Suficiencia y completitud

1. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable $X \to \{B(k,p);\ p \in 0,1\}$ y sea $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$. Probar, *a*) usando la definición y *b*) aplicando el teorema de factorización, que $T$ es suficiente para $p$.

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable $X \to \{\mathcal{P}(\lambda);\ \lambda \in \mathbb{R}^+\}$ y sea $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$. Probar, *a*) usando la definición y *b*) aplicando el teorema de factorización, que $T$ es suficiente para $\lambda$.

3. Sea $(X_1, X_2, X_3)$ una muestra aleatoria simple de una variable $X \to \{B(1,p);\ p \in (0,1)\}$. Probar que el estadístico $X_1 + 2X_2 + 3X_3$ no es suficiente.

4. Aplicando el teorema de factorización, y basándose en una muestra de tamaño arbitrario, encontrar un estadístico suficiente para cada una de las siguientes familias de distribuciones (en las familias biparamétricas, suponer los casos de sólo un parámetro desconocido y de los dos desconocidos).

   *a*) $X \to \{U(-\theta/2, \theta/2);\ \ \theta > 0\}$

   *b*) $X \to \{\Gamma(p,a);\ p > 0,\ a > 0\}$

   *c*) $X \to \{\beta(p,q);\ p > 0,\ q > 0\}$

   *d*) $X \to \{P_{N_1,N_2};\ N_1, N_2 \in \mathbb{N},\ N_1 \leq N_2\};\ P_{N_1,N_2}\{X = x\} = \dfrac{1}{N_2 - N_1 + 1},\ x = N_1, \ldots, N_2.$

5. Sea $X \to \{P_N;\ N \in \mathbb{N}\}$, siendo $P_N$ la distribución uniforme en los puntos $\{1, \ldots, N\}$, y sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X$. Probar que $\max(X_1, \ldots, X_n)$ es un estadístico suficiente y completo.

6. Basándose en una muestra de tamaño arbitrario, obtener un estadístico suficiente y completo para la familia de distribuciones definidas por todas las densidades de la forma
$$f_\theta(x) = e^{\theta - x}, \quad x > \theta.$$

7. Comprobar que las siguientes familias de distribuciones son exponenciales uniparamétricas y, considerando una muestra aleatoria simple de una variable con distribución en dicha familia, obtener, si existe, un estadístico suficiente y completo.

   *a*) $\{B(k_0, p);\ 0 < p < 1\}$

   *b*) $\{\mathcal{P}(\lambda);\ \lambda > 0\}$

   *c*) $\{BN(k_0, p);\ 0 < p < 1\}$

   *d*) $\{exp(\lambda);\ \lambda > 0\}$

8. Estudiar si las siguientes familias de distribuciones son exponenciales biparamétricas. En caso afirmativo, considerando una muestra aleatoria simple de una variable con distribución en dicha familia, obtener, si existe, un estadístico suficiente y completo.

    *a*) $\{\Gamma(p, a);\ p > 0,\ a > 0\}$

    *b*) $\{\beta(p, q);\ p > 0,\ q > 0\}$

**Tema 4: Estimación puntual. Insesgadez y mínima varianza**

1. Sea $(X_1, \ldots, X_n)$ una muestra de una variable $X \to \{\mathcal{N}(\mu, \sigma^2);\ \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$. Probar que
$$T(X_1, \ldots, X_n) = \begin{cases} 1 & \overline{X} \leq 0 \\ \\ 0 & \overline{X} > 0 \end{cases}$$
es un estimador insesgado de la función paramétrica $\Phi(-\mu\sqrt{n}/\sigma)$, siendo $\Phi$ la función de distribución de la $\mathcal{N}(0,1)$.

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \to \{B(1,p);\ p \in (0,1)\}$ y sea $T = \sum_{i=1}^{n} X_i$.

   *a*) Probar que si $k \in \mathbb{N}$ y $k \leq n$, el estadístico
   $$\frac{T(T-1)\cdots(T-k+1)}{n(n-1)\cdots(n-k+1)}$$
   es un estimador insesgado de $p^k$. ¿Es este estimador el UMVUE?

   *b*) Probar que si $k > n$, no existe ningún estimador insesgado para $p^k$.

   *c*) ¿Puede afirmarse que $\frac{T}{n}\left(1 - \frac{T}{n}\right)^2$ es insesgado para $p(1-p)^2$?

3. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable $X \to \{\mathcal{P}(\lambda);\ \lambda > 0\}$. Encontrar, si existe, el UMVUE para $\lambda^s$, siendo $s \in \mathbb{N}$ arbitrario.

4. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable con distribución uniforme discreta en los puntos $\{1, \ldots, N\}$, siendo $N$ un número natural arbitrario. Encontrar el UMVUE para $N$.

5. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$ cuya función de densidad es de la forma
$$f_\theta(x) = \frac{1}{2\sqrt{x\theta}}, \quad 0 < x < \theta$$
Calcular, si existe, el UMVUE para $\theta$.

6. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$ con función de densidad
$$f_\theta(x) = \frac{\theta}{x^2}, \quad x > \theta$$
Calcular, si existen, los UMVUE para $\theta$ y para $1/\theta$.

7. Sea $X \to \{P_\theta;\ \theta \in \mathbb{R}\}$ siendo $P_\theta$ una distribución con función de densidad

$$f_\theta(x) = e^{\theta - x}, \quad x \geq \theta.$$

Dada una muestra aleatoria simple de tamaño arbitrario, encontrar los UMVUE de $\theta$ y de $e^\theta$.

8. Sea $X$ la variable que describe el número de fracasos antes del primer éxito en una sucesión de pruebas de Bernoulli con probabilidad de éxito $\theta \in (0, 1)$, y sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X$.

   *a*) Probar que la familia de distribuciones de $X$ es regular y calcular la función de información asociada a la muestra.

   *b*) Especificar la clase de funciones paramétricas que admiten estimadores eficientes y los correspondientes estimadores.

   *c*) Calcular la varianza de cada estimador eficiente y comprobar que coincide con la correspondiente cota de Fréchet-Cramér-Rao.

   *d*) Calcular, si existen, los UMVUE para $P_\theta[X = 0]$ y para $E_\theta[X]$ y decir si son eficientes.

9. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$ con distribución exponencial.

   *a*) Probar que la familia de distribuciones de $X$ es regular.

   *b*) Encontrar la clase de funciones paramétricas que admiten estimador eficiente y el estimador correspondiente. Calcular la varianza de estos estimadores.

   *c*) Basándose en el apartado anterior, encontrar el UMVUE para la media de $X$.

   *d*) Dar la cota de Fréchet-Cramér-Rao para la varianza de estimadores insesgados y regulares de $\lambda^3$. ¿Es alcanzable dicha cota?

10. Sea $X$ una variable aleatoria con función de densidad de la forma

$$f_\theta(x) = \theta x^{\theta - 1}, \quad 0 < x < 1.$$

   *a*) Sabiendo que $\mathrm{E}_\theta[\ln X] = -\dfrac{1}{\theta}$ y $\mathrm{Var}_\theta[\ln X] = \dfrac{1}{\theta^2}$, comprobar que esta familia de distribuciones es regular.

   *b*) Basándose en una muestra aleatoria simple de $X$, dar la clase de funciones paramétricas con estimador eficiente, los estimadores y su varianza.

## Tema 5: Estimación de máxima verosimilitud y otros métodos

1. Sea $X \to \{P_\theta;\ \theta \in \mathbb{R}\}$ siendo $P_\theta$ una distribución con función de densidad
$$f_\theta(x) = e^{\theta - x}, \quad x \geq \theta.$$
Dada una muestra aleatoria simple de tamaño $n$, encontrar los estimadores máximo verosímiles de $\theta$ y de $e^\theta$. Basándose en los resultados del problema 7 de la relación 4, decir si estos estimadores son insesgados.

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X$ con distribución exponencial. Basándose en los resultados del problema 9 de la relación 4, encontrar los estimadores máximo verosímiles de la media y de la varianza de $X$.

3. Sea $X$ una variable aleatoria con función de densidad de la forma
$$f_\theta(x) = \theta x^{\theta - 1}, \quad 0 < x < 1$$
   *a*) Calcular un estimador máximo verosímil para $\theta$.
   *b*) Deducir dicho estimador a partir de los resultados del problema 10 de la relación 4.

4. Sea $(X_1, \cdots, X_n)$ una muestra de una variable $X \to \{B(k_0, p);\ p \in (0,1)\}$. Estimar, por máxima verosimilitud y por el método de los momentos, el parámetro $p$ y la varianza de $X$.

   *Aplicación*: Se lanza 10 veces un dado cargado y se cuenta el número de veces que sale un 4. Este experimento se realiza 100 veces de forma independiente, obteniéndose los siguientes resultados:

| $n^o$ de 4 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| frecuencia | 84 | 15 | 1 | 0 |

   Estimar, a partir de estos datos, la probabilidad de salir un cuatro.

5. Se lanza un dado hasta que salga un 4 y se anota el número de lanzamientos necesarios; este experimento se efectúa veinte veces de forma independiente. A partir de los resultados obtenidos, estimar la probabilidad de sacar un 4 por máxima verosimilitud.

6. En 20 días muy fríos, una granjera pudo arrancar su tractor en el primer, tercer, quinto, primer, segundo, primer, tercer, séptimo, segundo, cuarto, cuarto, octavo, primer, tercer, sexto, quinto, segundo, primer, sexto y segundo intento. Suponiendo que la probabilidad de arrancar en cada intento es constante, y que las observaciones se han obtenido de forma independiente, dar la estimación más verosímil de la probabilidad de que el tractor arranque en el segundo intento.

7. Una variable aleatoria discreta toma los valores 0,1 y 2 con las siguientes probabilidades

$$P_p(X=0)=p^2, \qquad P_p(X=1)=2p(1-p), \qquad P_p(X=2)=(1-p)^2,$$

siendo $p$ un parámetro desconocido. En una muestra aleatoria simple de tamaño 100, se ha presentado 22 veces el 0, 53 veces el 1 y 25 veces el 2. Calcular la función de verosimilitud asociada a dicha muestra y dar la estimación más verosímil de $p$.

8. En el muestreo de una variable aleatoria con distribución $\mathcal{N}(\mu,1)$, $\mu \in \mathbb{R}$, se observa que no se obtiene un valor menor que -1 hasta la quinta observación. Dar una estimación máximo verosímil de $\mu$.

9. En la producción de filamentos eléctricos la medida de interés, $X$, es el tiempo de vida de cada filamento, que tiene una distribución exponencial de parámetro $\theta$. Se eligen $n$ de tales filamentos de forma aleatoria e independiente, pero, por razones de economía, no conviene esperar a que todos se quemen y la observación acaba en el tiempo $T$. Dar el estimador máximo verosímil para la media de $X$ a partir del número de filamentos quemados durante el tiempo de observación.

10. Sean $X_1, \cdots, X_n$ observaciones independientes de una variable $X \to \{\Gamma(p,a);\ p,a>0\}$. Estimar ambos parámetros mediante el método de los momentos.

    *Aplicación*: Ciertos neumáticos radiales tuvieron vidas útiles de 35200, 41000, 44700, 38600 y 41500 kilómetros. Suponiendo que estos datos son observaciones independientes de una variable con distribución exponencial de parámetro $\theta$, dar una estimación de dicho parámetro por el método de los momentos.

## Tema 6: Estimación por intervalos de confianza

1. Sea $\overline{X}$ la media de una muestra aleatoria de tamaño $n$ de una población $\mathcal{N}(\mu, 16)$. Encontrar el menor valor de $n$ para que $(\overline{X} - 1,\ \overline{X} + 1)$ sea un intervalo de confianza para $\mu$ al nivel de confianza 0.9.

2. La altura en cm. de los individuos varones de una población sigue una distribución $\mathcal{N}(\mu, 56.25)$. Si en una muestra aleatoria simple de tamaño 12 de dicha población se obtiene una altura media de 175 cm., determinar un intervalo de confianza para $\mu$ al nivel de confianza 0.95. ¿Qué tamaño de muestra es necesario para que el intervalo de confianza a dicho nivel tenga longitud menor que 1 cm?

3. Una fábrica produce tornillos cuyo diámetro medio es 3 mm. Se seleccionan aleatoria e independientemente 12 de estos tornillos y se miden sus diámetros, que resultan ser 3.01, 3.05, 2.99, 2.99, 3.00, 3.02, 2.98, 2.99, 2.97, 2.97, 3.02 y 3.01. Suponiendo que el diámetro es una variable aleatoria con distribución normal, determinar un intervalo de confianza para la varianza al nivel de confianza 0.99, y una cota superior de confianza al mismo nivel. Interpretar los resultados en términos de la desviación típica del diámetro de los tornillos.

4. Las notas en cierta asignatura de 7 alumnos de una clase, elegidos de forma aleatoria e independiente son: 4.5, 3, 6, 7, 1.5, 5.2 y 3.6. Suponiendo que las notas tienen distribución normal, dar un intervalo de confianza para la varianza de las mismas al nivel de confianza 0.95.

5. Dos muestras independientes, cada una de tamaño 7, de poblaciones normales con igual varianza, producen medias 4.8 y 5.4 y cuasivarianzas muestrales 8.38 y 7.62, respectivamente. Encontrar un intervalo de confianza para la diferencia de medias al nivel de confianza 0.95.

6. La siguiente tabla presenta los salarios anuales (en miles de euros) de dos grupos de recién graduados de dos carreras diferentes. Suponiendo normalidad en los salarios de ambos grupos, determinar un intervalo de confianza para el cociente de las varianzas al nivel de confianza 0.90.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GRUPO 1 | 16.3 | 18.2 | 17.5 | 16.1 | 15.9 | 15.4 | 15.8 | 17.3 | 14.9 | 15.1 | | | | |
| GRUPO 2 | 13.2 | 15.1 | 13.9 | 14.7 | 15.6 | 15.8 | 14.9 | 18.1 | 15.6 | 15.3 | 16.2 | 15.2 | 15.4 | 16.6 |

7. Con objeto de estudiar la efectividad de un agente diurético, se eligen al azar 11 pacientes, aplicando dicho fármaco a seis de ellos y un placebo a los cinco restantes. La variable observada en esta experiencia fue la concentración de sodio en la orina a las 24 horas, que se supone tiene una distribución normal en ambos casos. Los resultados observados fueron:

   DIURÉTICO: 20.4, 62.5, 61.3, 44.2, 11.1, 23.7

   PLACEBO: 1.2, 6.9, 38.7, 20.4, 17,2

*a*) Calcular un intervalo de confianza para el cociente de las varianzas al nivel de confianza 0.95.

*b*) Suponiendo que las varianzas son iguales, calcular un intervalo de confianza para la diferencia de las medias al nivel de confianza 0.9, y una cota inferior de confianza al mismo nivel. Interpretar los resultados.

8. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria con distribución $U(0, \theta)$. Dado un nivel de confianza arbitrario, calcular el intervalo de confianza para $\theta$ de menor longitud media uniformemente basado en un estadístico suficiente.

9. Utilizando la desigualdad de Chebychev, dar un intervalo de confianza para $p$ a nivel de confianza arbitrario, basado en una muestra de tamaño arbitrario de una variable aleatoria con distribución $B(1, p)$.

10. Para una muestra de tamaño $n$ de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{2x}{\theta^2}, \quad 0 < x < \theta,$$

encontrar el intervalo de confianza para $\theta$ de menor longitud media uniformemente a nivel de confianza $1 - \alpha$, basado en un estadístico suficiente.

11. Para una muestra de tamaño $n$ de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{\theta}{x^2}, \quad x > \theta$$

encontrar el intervalo de confianza para $\theta$ de menor longitud media uniformemente a nivel de confianza $1 - \alpha$, basado en el estimador máximo verosímil de $\theta$.

## Tema 8: Regresión lineal y análisis de la varianza

1. Una compañía de energía eléctrica pretende desarrollar un modelo lineal para el consumo de energía en función de la temperatura media diaria durante los meses de invierno. En 9 días elegidos al azar se obtuvo la siguiente información:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Temperatura (º C) | 0 | 2 | 4 | 8 | 13 | -4 | -6 | -8 | -11 |
| Consumo (Kw/día) | 70 | 79 | 67 | 66 | 63 | 97 | 82 | 90 | 107 |

*a*) Obtener la recta de regresión estimada e interpretar sus coeficientes.

*b*) Descomponer la variabilidad de los datos de consumo. Obtener la varianza residual, el coeficiente de determinación y dar su interpretación.

*c*) Obtener la predicción del consumo de energía para un día con temperatura media 10 grados y estimar el error cuadrático medio de esta predicción.

*d*) Si se suponen las hipótesis adecuadas de normalidad, ¿se puede considerar, al nivel de significación $\alpha = 0.05$, que el consumo no depende linealmente de la temperatura?

2. Una compañía de seguros desea establecer una relación lineal, y el grado de dicha relación, para determinar el montante anual del seguro de vida del cabeza de familia en función del ingreso mensual familiar. Observada una muestra aleatoria de 10 familias elegidas de forma independiente, se obtuvo la siguiente información:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Ingreso (cientos de euros) | 10 | 10 | 15 | 20 | 20 | 25 | 25 | 30 | 30 | 30 |
| Seguro (decenas de euros) | 50 | 35 | 55 | 55 | 70 | 65 | 65 | 60 | 75 | 90 |

*a*) ¿Cabe pensar en la citada relación lineal? En caso afirmativo, estimar la recta de regresión, interpretar los coeficientes y obtener la predicción del montante del seguro de vida para un ingreso mensual de 2800 euros.

*b*) Suponiendo las hipótesis adecuadas de normalidad, realizar el contraste de regresión al nivel de significación $\alpha = 0.05$ e interpretar el resultado.

3. Los datos de la siguiente tabla representan las calificaciones medias ($X$) de 7 recién graduados y sus respectivos salarios iniciales ($Y$) en miles de euros:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X$ | 2 .75 | 2.85 | 2.95 | 3.05 | 3.2 | 3.4 | 3.6 |
| $Y$ | 16.8 | 18.8 | 17.2 | 17.2 | 21.2 | 21.5 | 22.4 |

*a*) Estimar la recta de regresión de $Y$ sobre $X$ e interpretar sus coeficientes.

*b*) Determinar la varianza residual.

*c*) Calcular e interpretar el coeficiente de determinación.

*d*) Predecir el salario inicial de un estudiante con una calificación media de 3.25.

*e*) Suponiendo las hipótesis adecuadas de normalidad, contrastar la hipótesis de que no existe relación lineal entre ambas variables al nivel de significación $\alpha = 0.01$.

4. En cierto estudio sobre la relación entre el diámetro de los guisantes ($X$) y el diámetro medio de sus descendientes ($Y$), Galton obtuvo los siguientes resultados:

| D. Padres | 21 | 20 | 19 | 18 | 17 | 16 | 15 |
|---|---|---|---|---|---|---|---|
| D. Descendientes | 17.26 | 17.07 | 16.37 | 16.4 | 16.13 | 16.17 | 15.98 |

*a*) Determinar el modelo de regresión lineal estimado de $Y$ sobre $X$ e interpretar el valor estimado de la pendiente. Dar la predicción del diámetro de los guisantes cuyos progenitores tienen un diámetro de 18.5. Dar una medida de la bondad del ajuste de los datos a la recta estimada.

*b*) Suponiendo las hipótesis adecuadas de normalidad, ¿puede deducirse, al nivel 0.05, que no hay relación lineal entre las variables consideradas? Relacionar este resultado con las conclusiones anteriores.

5. Una compañía farmacéutica investiga los efectos de 5 compuestos. El experimento consiste en inyectar los compuestos a 13 ratas de características similares y anotar los tiempos de reacción. Los animales se clasifican en 5 grupos de 4, 2, 2, 3 y 2 ratas, respectivamente, y a cada grupo se le administra un compuesto diferente, obteniéndose los resultados de la siguiente tabla:

| Grupo | Tiempo de reacción (minutos) | | | |
|---|---|---|---|---|
| 1 | 8.3 | 7.6 | 8.4 | 8.3 |
| 2 | 7.4 | 7.1 | | |
| 3 | 8.1 | 6.4 | | |
| 4 | 7.9 | 8.5 | 10.0 | |
| 5 | 7.1 | 8 | | |

Suponiendo que se verifican las hipótesis de normalidad, aleatoriedad, independencia e igualdad de varianzas, contrastar la hipótesis de que los tiempos medios de reacción coinciden en los cinco grupos y, por tanto, la eficacia de los cinco compuestos es la misma.

6. Se quiere estudiar la eficacia de tres fertilizantes, $A, B$ y $C$, en la producción de cierto fruto. Para ello se aplica el $A$ en 8 parcelas, el $B$ en 6, y el $C$ en 12 parcelas. Las parcelas son de características similares en cuanto a fertilidad, por lo que se considera que las diferencias en la producción, si las hay, serán debidas al tipo de fertilizante. Las toneladas producidas en cada parcela en una determinada temporada son:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 6 | 7 | 5 | 6 | 5 | 8 | 4 | 7 | | | | |
| $B$ | 10 | 9 | 9 | 10 | 10 | 6 | | | | | | |
| $C$ | 3 | 4 | 8 | 3 | 7 | 6 | 3 | 6 | 4 | 7 | 6 | 3 |

Suponiendo que las tres muestras proceden de poblaciones normales con varianzas iguales, contrastar la hipótesis de que los abonos son igualmente eficaces.

7. Los siguientes datos corresponden a observaciones del consumo medio (en Kw/h) realizado por 5 tipos de calefactores para mantener una habitación a una temperatura determinada durante todo un día:

| Tipo | Consumo (en kw/h) | | | | | |
|---|---|---|---|---|---|---|
| 1 | 14.5 | 14.1 | 14.6 | 14.2 | | |
| 2 | 13.2 | 13.4 | 13.0 | | | |
| 3 | 13.7 | 13.6 | 14.1 | 13.8 | 14.0 | |
| 4 | 12.7 | 13.1 | 12.8 | 12.9 | 13.3 | 13.2 |
| 5 | 14.6 | 15.2 | 14.4 | 14.8 | 14.3 | |

Contrastar la hipótesis de igualdad de los consumos medios de los diferentes tipos de calefactores. ¿Bajo qué hipótesis se puede realizar este contraste?

8. En un tratamiento contra la hipertensión se seleccionaron 35 enfermos de características similares. Los enfermos se distribuyeron en cuatro grupos de 10 (P, A, B y AB). El grupo P tomó "placebo"(fármaco inocuo), el grupo A tomó un fármaco "A", el grupo B un fármaco "B" y el grupo AB una asociación entre "A" y "B". Para valorar la eficacia de los tratamientos, se registró el descenso de la presión diastólica desde el inicio del tratamiento hasta después de una semana de tratamiento. Los resultados, después de registrarse algunos abandonos, fueron:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| P | 10 | 0 | 15 | -20 | 0 | 15 | -5 | | | |
| A | 20 | 25 | 33 | 25 | 30 | 18 | 27 | 0 | 35 | 20 |
| B | 15 | 10 | 25 | 30 | 15 | 35 | 25 | 22 | 11 | 25 |
| AB | 10 | 5 | -5 | 15 | 20 | 20 | 0 | 10 | | |

A la vista de estos datos, ¿Puede afirmarse que el descenso de la presión diastólica coincide en los cuatro grupos? ¿Bajo qué hipótesis?

**Tema 9: Contrastes de hipótesis no paramétricos**

1. A partir de los siguientes datos, que muestran el número de accidentes en un determinado regimiento del ejército durante 200 días elegidos al azar, contrastar si el número de accidentes diarios sigue una distribución de Poisson de parámetro 2.

| *Nº de accidentes* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| *Nº de días* | 22 | 53 | 58 | 39 | 20 | 5 | 2 | 1 |

2. Una tela cuadrada tiene 60 defectos de fabricación. Con objeto de analizar la distribución de los defectos en la superficie de la tela, se ha dividido ésta en 9 zonas cuadradas exactamente iguales, observándose los siguientes defectos en cada zona:

| | | |
|---|---|---|
| 8 | 7 | 3 |
| 5 | 9 | 11 |
| 6 | 4 | 7 |

Contrastar, a partir de estos datos, si los defectos se distribuyen uniformemente en toda la superficie o, por el contrario, siguen algún patrón de ocurrencia.

3. Un modelo genético indica que la distribución de una población de hombres y mujeres, daltónicos o no, se ajusta a probabilidades de la forma:

| | *Hombres* | *Mujeres* |
|---|---|---|
| *No daltónicos* | $(1-p)/2$ | $(1-p^2)/2$ |
| *Daltónicos* | $p/2$ | $p^2/2$ |

Para comprobar esta teoría se examinaron 2000 individuos de la población, elegidos al azar, obteniéndose los siguientes resultados:

| | *Hombres* | *Mujeres* |
|---|---|---|
| *No daltónicos* | 894 | 1015 |
| *Daltónicos* | 81 | 10 |

Contrastar, mediante el test de la $\chi^2$, si esta muestra concuerda con el modelo teórico.

4. Un laboratorio farmacéutico afirma que uno de sus productos confiere inmunidad a la picadura de insectos durante un tiempo exponencial de media 2.5 horas. Probado el producto en 20 sujetos, en un ambiente con gran número de mosquitos, los instantes (en horas) en que recibieron la primera picadura fueron:

0.01, 0.02 0.03, 0.23, 0.51, 0.74, 0.96, 1.17, 1.46, 1.62,
2.18, 2.25, 2.79, 3.45, 3.83, 3.92, 4.27, 5.43, 5.79, 6.34.

Usando el test de Kolmogorov-Smirnov, contrastar, a partir de estos datos, si puede aceptarse la afirmación del laboratorio.

5. Cierta comunidad ha modificado la procedencia del agua destinada al consumo doméstico. Se sabe que, con el antiguo suministro, la distribución de la cantidad de sodio por unidad de volumen de sangre de sus habitantes es simétrica alrededor de 3.24 gr. Tras cierto tiempo, se quiere comprobar si la modificación ha afectado a la concentración de sodio, en el sentido de que su distribución se haya trasladado o no. Para ello, se han realizado 15 análisis con los siguientes resultados (en gr. por unidad):

   2.37 2.95 3.4 2.64 3.66 3.18 2.72 3.61 3.87 1.97 1.66 3.72 2.10 1.83 3.03

   ¿Se puede afirmar, al nivel de significación 0.1, que la distribución de la cantidad de sodio no ha variado?

6. La siguiente tabla presenta las presiones sanguíneas sistólicas de 10 individuos antes y después de haber dejado la bebida.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 140 | 165 | 160 | 160 | 175 | 190 | 170 | 175 | 155 | 160 |
| $D$ | 145 | 150 | 150 | 160 | 170 | 175 | 160 | 165 | 145 | 170 |

   ¿Se puede afirmar a partir de los datos que el abandono de la bebida no disminuye la presión sanguínea? ¿Bajo qué hipótesis?

7. En cierta comunidad de E.E.U.U. se realizó un estudio para investigar si el sueldo anual de las familias influía en los hijos para la elección de los diferentes cursos de enseñanza secundaria (Preparatorio, General y Comercial). Para ello, se hizo una clasificación de los sueldos en cuatro niveles (I, II, III y IV), y se tomó una muestra aleatoria simple de 390 estudiantes, obteniéndose la siguiente tabla de frecuencias:

| *Sueldo* / *Curso* | I | II | III | IV |
|---|---|---|---|---|
| Preparatorio | 23 | 40 | 16 | 2 |
| General | 11 | 75 | 107 | 14 |
| Comercial | 1 | 31 | 60 | 10 |

   A la vista de los datos, decidir, al nivel de significación 0.01, si se acepta que el nivel económico familiar no influye en la decisión de los estudiantes a la hora de elegir curso.

8. En un estudio sociológico sobre la polución atmosférica se entrevistó a 40 residentes de cada una de tres zonas residenciales en Gran Bretaña. La siguiente tabla muestra las respuestas a la pregunta: ¿Hay problema de polución en su barrio?

| *Respuesta* / *Zona residencial* | No | Sí | No sabe | No contesta |
|---|---|---|---|---|
| 1 | 5 | 31 | 2 | 2 |
| 2 | 10 | 21 | 4 | 5 |
| 3 | 11 | 20 | 7 | 2 |

   Contrastar si las tres poblaciones de residentes pueden considerarse homogéneas con respecto a su opinión sobre la polución.

9. Para determinar si diferentes tipos de profesiones de los individuos activos de cierto colectivo afectan a la tensión arterial, se clasificó a los individuos en cuatro grupos, atendiendo a su profesión, y se midió la tensión a una muestra de individuos elegidos de forma aleatoria. Clasificando la tensión en los niveles "Bajo", "Normal" y "Alto", se obtuvo los siguientes resultados, que muestran el número de individuos de cada tipo de profesión con los distintos niveles de tensión:

| *Profesión* \ *Nivel de tensión* | Bajo | Normal | Alto |
|---|---|---|---|
| I | 8 | 4 | 3 |
| II | 5 | 7 | 7 |
| III | 4 | 8 | 8 |
| IV | 5 | 7 | 8 |

¿Qué conclusión acerca del problema planteado se obtiene a la vista de estos datos? Especificar las hipótesis nula y alternativa que se contrastan.

10. Para determinar si las calificaciones de los alumnos en selectividad son independientes de las calificaciones en bachiller, se eligió de forma aleatoria una muestra de alumnos, a los que se preguntó ambas calificaciones, obteniendo los siguientes resultados:

| *Bachiller* \ *Selectividad* | Suspenso | Aprobado | Notable | Sobresaliente |
|---|---|---|---|---|
| Aprobado | 10 | 6 | 4 | 5 |
| Notable | 7 | 9 | 9 | 4 |
| Sobresaliente | 6 | 10 | 10 | 6 |

¿Qué conclusión acerca del problema planteado se obtiene a la vista de estos datos? Especificar las hipótesis nula y alternativa que se contrastan.

## Tema 7: Contraste de hipótesis

1. Se toma una observación de una variable con distribución de Poisson para contrastar que la media vale 1 frente a que vale 2.

   *a*) Construir un test no aleatorizado con nivel de significación 0.05 para el contraste planteado. Calcular las probabilidades de cometer error de tipo 1 y de tipo 2, el tamaño y la potencia del test frente a la hipótesis alternativa.

   *b*) ¿Cómo debe aleatorizarse el test para alcanzar el tamaño 0.05? ¿Cuál es la potencia de este test?

2. Una urna contiene 10 bolas, blancas y negras. Para contrastar que el número de bolas blancas es 5 frente a que dicho número es 6 o 7, se extraen tres bolas con reemplazamiento y se rechaza $H_0$ sólo si se obtienen 2 o 3 bolas blancas. Calcular el tamaño de este test y la potencia frente a las alternativas.

3. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria con distribución de Poisson de parámetro $\lambda$. Encontrar el test más potente de tamaño $\alpha$ para resolver el problema de contraste

$$\begin{array}{c} H_0 : \lambda = \lambda_0 \\ H_1 : \lambda = \lambda_1 \end{array}$$

   *Aplicación:* En una centralita telefónica el número de llamadas por minuto sigue una distribución de Poisson. Si en cinco minutos se han recibido 12 llamadas, ¿puede aceptarse que el número medio de llamadas por minuto es 1.5, frente a que dicho número es 2, al nivel de significación 0.05? Calcular la potencia del test obtenido.

4. Dada una muestra de tamaño $n$ de una variable con distribución $\mathcal{N}(\mu, \sigma_0^2)$, deducir el test más potente de tamaño arbitrario para contrastar hipótesis simples sobre $\mu$.

5. Dada una muestra de tamaño $n$ de una variable aleatoria con distribución $U(-\theta, \theta)$, deducir el test más potente de tamaño $\alpha$ para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ y calcular su potencia. ¿Cuál es el test óptimo fijado un nivel de significación arbitrario?

6. Deducir el test más potente de tamaño arbitrario para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, basándose en una muestra de tamaño $n$ de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{\theta}{x^2}, \quad x > \theta.$$

   Deducir el test óptimo para un nivel de significación arbitrario.

7. Construir el test de Neyman-Pearson de tamaño $\alpha$ para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, basándose en una muestra de tamaño $n$ de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{1}{x \ln \theta}, \quad 1 < x < \theta.$$

   Deducir el test óptimo para un nivel de significación arbitrario.

8. Sea $X$ una observación de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{1}{\theta} \mathrm{e}^{-x/\theta}, \quad x > 0.$$

Construir el test de razón de verosimilitudes de tamaño $\alpha$ arbitrario para contrastar

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta < \theta_0.$$

9. En base a una observación de $X \to \{B(n,p);\ p \in (0,1)\}$, deducir el test de razón de verosimilitudes para contrastar la hipótesis de que el parámetro $p$ no supera un determinado valor, $p_0$.

10. Sea $X$ una variable con función de densidad

$$f_\theta(x) = \theta x^{\theta-1}, \quad 0 < x < 1.$$

Basándose en una observación de $X$, deducir el test de razón de verosimilitudes de tamaño arbitrario para contrastar

$$H_0 : \theta \leq \theta_0$$
$$H_1 : \theta > \theta_0.$$

11. Un fabricante de coches asegura que la distancia media recorrida con un galón de gasolina es al menos 30 millas. Probados 9 coches de esta fábrica, la distancia media recorrida con un galón de gasolina ha sido 26 millas, y la suma de los cuadrados 6106 millas al cuadrado.

    *a*) Suponiendo que la distancia recorrida por estos coches con un galón de gasolina tiene distribución normal, contrastar la hipótesis del fabricante a partir de estos datos, a nivel de significación 0.01.

    *a*) ¿Qué conclusión se obtendría de estos mismos datos, al mismo nivel de significación, si se sabe que la desviación típica de la variable considerada es 5.5?

12. Un fabricante de baterías asegura que la desviación típica del tiempo de vida de las mismas es, a lo sumo, 70 horas. Una muestra de 26 baterías tomadas al azar ha dado una cuasidesviación típica de 84 horas. Haciendo las hipótesis adecuadas de normalidad, ¿proporcionan los datos evidencia para rechazar la hipótesis del fabricante al nivel 0.02?

13. Un profesor asegura que tiene un nuevo método de enseñanza mejor que el usado tradicionalmente. Para comprobar si tiene razón se selecciona de forma aleatoria e independiente dos grupos de alumnos, $A$ y $B$, utilizándose el nuevo método con el grupo $A$ y el tradicional con el $B$. A final de curso se hace un examen a los alumnos, obteniéndose las siguientes puntuaciones:

    Grupo $A$: 6, 5, 4, 7, 3, 5.5, 6, 7, 6

Grupo $B$: 5, 4, 5, 6, 4, 6, 5, 3, 7

Supuesto que las puntuaciones de cada grupo tienen distribución normal, ¿proporcionan estos datos evidencia para rechazar el nuevo método a nivel de significación 0.05?

14. A partir de las siguientes observaciones de muestras independientes de dos poblaciones normales, contrastar, al nivel de significación 0.01, si la media de la primera población supera en al menos una unidad la media de la segunda.

    MUESTRA 1: 132, 139, 126, 114, 122, 132, 141, 126.

    MUESTRA 2: 124, 141, 118, 116, 114, 132, 145, 123, 121.

15. Una central lechera recibe diariamente leche de dos granjas A y B. Para comparar la calidad de los productos recibidos se ha medido el contenido en grasa en muestras de leche tomadas al azar de cada una de las granjas, con los siguientes resultados:

| | Contenido en grasa (%) | | | | | |
|---|---|---|---|---|---|---|
| Granja A | 14 | 12 | 15 | 15 | 11 | 16 |
| Granja B | 20 | 18 | 18 | 19 | 15 | |

   *a*) ¿Puede suponerse, a nivel de significación 0.05, que el contenido medio en grasa de la leche de las dos granjas es el mismo? Especificar las hipótesis bajo las que se resuelve este problema.

   *b*) Calcular un intervalo de confianza, a nivel de confianza 0.9, para la varianza del contenido en grasa de la leche de la granja B. A partir de dicho intervalo, deducir si puede aceptarse que la varianza de esta población es igual a 3. Especificar el problema de contraste y el test utilizado; calcular su tamaño.