EJERCICIOS DE DISTRIBUCIONES MUESTRALES

(1º) Sea $\bar{y}_{25}$ la media de una muestra aleatoria de tamaño n = 25 extraída de una distribución de probabilidad con densidad desconocida f(y), media μ = 17 y desviación estándar σ = 10. De forma similar, sea $\bar{y}_{100}$ la media de una muestra aleatoria de tamaño n = 100 seleccionada de la misma distribución de probabilidad.

(a) Describe las distribuciones de muestreo $\bar{y}_{25}$ y $\bar{y}_{100}$.

(b) ¿Cuál de las probabilidades, p(15 < $\bar{y}_{25}$ < 19) o p(15 < $\bar{y}_{100}$ < 19),esperas que sea mayor?

(c) Calcula aproximaciones de tales probabilidades.

(2º) El Centro de Apoyo de Ingeniería y Vivienda del Ejército de Estados Unidos patrocinó un estudio de las características de confiabilidad, disponibilidad y mantenibilidad de sistemas pequeños que trabajan con diesel y gas en instalaciones comerciales y militares (IEEE Tran. on Industry Applications, julio/agosto, 1990). El estudio reveló que el tiempo Y antes de que sea necesario dar mantenimiento correctivo a sistemas diesel auxiliares continuos tiene una distribución exponencial aproximada con una media estimada en 1700 horas.

(a) Suponiendo μ = 1700, calcula la probabilidad de que el tiempo medio antes de dar mantenimiento correctivo a una muestra de 70 sistemas diesel auxiliares continuos exceda las 2500 horas.

(b) Si observas $\bar{y} > \mathbf{2500}$, ¿qué conclusión harías acerca de μ?.

(3º) Estudios realizados por neurocientíficos del MIT revelan que la melatonina, segregada por la glándula pineal en el cerebro, funciona naturalmente como hormona inductora del sueño (Tampa Tribune, 1 de marzo de 1994). Voluntarios de sexo masculino recibieron distintas dosis de melanina o placebos y luego se colocaron en una habitación oscura a medio día, pidiéndoseles que cerraran los ojos y se durmieran. Lo que interesaba a los científicos del MIT era el tiempo Y (en minutos) que tardaba cada voluntario en quedarse dormido. Los investigadores determinaron que con el placebo (es decir, sin hormona), el tiempo medio para dormirse era de 15 minutos. Supón que con el tratamiento de placebo μ = 15 y σ = 5.

(a) Considera una muestra aleatoria de n = 20 hombres que reciben la hormona inductora del sueño. Sea $\bar{y}$ el tiempo medio en quedarse dormido en esta muestra. Si la hormona no es eficaz para inducir el sueño, describe la distribución de muestreo de $\bar{y}$.

(b) Calcula en el caso (a) p($\bar{y} \leq \mathbf{6}$).

(c) En el estudio real, el tiempo medio para quedarse dormidos de los 20 voluntarios fue $\bar{y}$ =5. Utiliza este resultado para hacer una inferencia acerca del verdadero valor de μ para aquellos que tomaron la melatonina.

(4º) Estimación de costos es el término con que se describe el proceso mediante el cual los ingenieros estiman el coste de los contratos de trabajo (por ejemplo, construcción de carreteras, construcción de edificios, etc.) que se han de otorgar a quien ofrezca el precio más bajo. La estimación de los ingenieros es la línea base con la cual se compra la licitación más baja (ganadora). En un estudio se investigaron los factores que afectan la exactitud de las estimaciones de los ingenieros (Cost Engineering, octubre de 1988), considerando la exactitud como el porcentaje de diferencia entre la licitación más baja y la estimación de los ingenieros. Uno de los factores más importantes es el número de licitadores: cuanto mayor sea el número de licitadores interesados en el contrato, más probable será que los ingenieros se excedan en la estimación del coste. En el caso de los contratos de construcción con cinco licitadores, el porcentaje promedio de error fue de –7.02 y la desviación estándar fue de 24.66. Considera una muestra de 50 contratos de construcción de edificios, cada uno con cinco licitadores.

(a) Describe la distribución de muestreo de $\bar{y}$, el porcentaje medio de diferencia entre la licitación baja y la estimación de los ingenieros, para los 50 contratos.
(b) Calcula p($\bar{y} < 0$) (ésta es la probabilidad de excederse en la estimación).
(c) Supón que se observa $\bar{y}$=-17.83 para una muestra de 50 contratos de construcción de edificios. Con base en la información proporcionada, ¿es verosímil que todos estos contratos tengan cinco licitadores?. Explica la respuesta.

(5º) Muchas especies de ranas terrestres que hibernan en la superficie del suelo o cerca de ella pueden sobrevivir a una exposición prolongada a las temperaturas bajas del invierno. En condiciones de congelación, la temperatura del cuerpo de la rana (temperatura de sobreenfriamiento) se mantiene relativamente alta gracias a una acumulación de glicerina en los fluidos corporales. Ciertos estudios han revelado que la temperatura de sobreenfriamiento de ranas terrestres congeladas a –6ºC tiene una distribución de frecuencia relativa con una media de –2.18ºc y una desviación estándar de 0.32ºC (Science, mayo de 1983). Considera la temperatura de sobreenfriamiento media $\bar{y}$ de una muestra aleatoria de n = 42 ranas terrestres congeladas a –6ºC.

(a) Calcula la probabilidad de que $\bar{y}$ sea mayor que –2.05ºC.
(b) Calcula la probabilidad de que $\bar{y}$ esté entre –2,20ºC y –2,10ºC.

(6º) Un artículo en Industrial Engineering (agosto de 1990) analizó la importancia de modelar el tiempo de inactividad de máquinas correctamente en los estudios de simulación. Como ilustración, el investigador consideró un sistema de una sola máquina herramienta con tiempos de reparación (en minutos) que se pueden modelar mediante una distribución gamma con parámetros $\alpha = 1$ y $\beta = 60$. Lo que interesa es el tiempo de reparación medio $\bar{y}$ de una muestra de 100 descomposturas de máquinas.

(a) Calcula E($\bar{y}$) y σ^2($\bar{y}$).

(b) ¿Qué distribución de probabilidad es el mejor modelo para la distribución de muestreo $\bar{y}$?. ¿Por qué?.

(c) Calcula la probabilidad de que el tiempo medio de reparación $\bar{y}$ no sea mayor que 30 minutos.

(7º) Supón que $\bar{y}$ y s^2 son la media y la varianza de una muestra de n observaciones de una población distribuida normalmente con media μ y varianza σ^2. Puede demostrarse que $\bar{y}$ y s^2 son estadísticamente independientes si la población maestreada tiene una distribución normal. Utiliza este resultado para demostrar que

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

tiene una distribución t con $\nu = (n-1)$ grados de libertad.

(8º) Sean $y_1, y_2, ..., y_n$, una muestra aleatoria de n observaciones de una distribución normal con media μ y varianza σ^2. Sea s^2 la varianza de la muestra. Utiliza las tablas de la distribución ji cuadrada para estimar las siguientes probabilidades:

(a) $p(s^2 > 8)$ cuando $n = 10$ y $\sigma^2 = 5$
(b) $p(s^2 > 1.11)$ cuando $n = 5$ y $\sigma^2 = 0.3$
(c) $p(s^2 > 199)$ cuando $n = 22$ y $\sigma^2 = 107$

(9º) En IEEE Transactions (junio de 1990) se presentó un algoritmo híbrido para resolver de programación matemáticos polinomiales 0-1. El tiempo de resolución (en segundos) de un problema escogido al azar y resuelto empleando el algoritmo tiene una distribución de probabilidad normal con media $\mu = 0.8$ segundos y varianza $\sigma = 1.5$ segundos. Considera una muestra aleatoria de $n = 30$ problemas resueltos con el algoritmo híbrido.

(a) Describe la distribución de muestreo s^2, la varianza de los tiempos de resolución para los 30 problemas.
(b) Determina la probabilidad aproximada de que s^2 sea mayor que 3.30.

(10º) Sean $y_1, y_2, ..., y_n$, una muestra aleatoria de n_1 observaciones de una distribución normal con media μ_1 y varianza σ_1^2. Sean $x_1, x_2, ..., x_n$, una muestra aleatoria de n_2 observaciones de una distribución normal con media μ_2 y varianza σ_2^2. Suponiendo que las muestras se seleccionaron de forma independiente, demuestra que

$$F = \frac{s_1^2}{s_2^2} \frac{\sigma_2^2}{\sigma_1^2}$$

tiene una distribución F con $\nu_1 = (n_1-1)$ grados de libertad en el numerador y $\nu_2 = (n_2-1)$ grados de libertad en el denominador.

(11º) Sean s_1^2 y s_2^2 las varianzas de muestras aleatorias independientes de tamaños n_1 y n_2 seleccionadas de poblaciones distribuidas normalmente con parámetros (μ_1, σ^2) y (μ_2, σ^2), respectivamente. O sea que las poblaciones tienen diferentes medias pero una varianza común σ^2. A fin de estimar la varianza común, se puede combinar información de ambas muestras y utilizar el estimador conjunto

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

Utiliza los teoremas que se han visto en el tema para demostrar que $(n_1 + n_2 - 2)\dfrac{s^2}{\sigma^2}$ tiene una distribución ji cuadrada con $\nu = (n_1+n_2-2)$ grados de libertad.

(12º) Sean $\bar{y}_1$ y $\bar{y}_2$ las medias de muestras aleatorias independientes con tamaños n_1 y n_2 seleccionadas de poblaciones distribuidas normalmente con parámetros (μ_1, σ^2) y (μ_2, σ^2), respectivamente. Si

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

demuestra que

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

tiene una distribución t de Student con $\nu = (n_1+n_2-2)$ grados de libertad.

(13º) Se dice que una variables aleatoria continua Y tiene una distribución lognormal con parámetros μ y σ si su función de densidad de probabilidad f(y) satisface

$$f(y) = \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \qquad (y > 0)$$

Demuestra que x = ln(y) tiene una distribución normal con media μ y varianza σ^2.

(14º) La dioxina, compuesto químico que muchas veces ha sido descrito como el más tóxico que se conoce, se crea como subproducto de la fabricación de herbicidas como el Agente Naranja. Los científicos han comprobado que 0.000005 gramos (cinco millonésimas de gramo) de dioxina –un puntito apenas visible para el ojo humano- constituyen una dosis letal para cobayas experimentales en más de la mitad de los animales ensayados, lo que hace a la dioxina 2000 veces más tóxica que la estricnina. Supón que la cantidad de dioxina requerida para matar un cobaya tiene una distribución de frecuencia relativa con media μ = 0.000005 gramos y desviación estándar σ = 0.000002 gramos. Considera un experimento en el que se mide la cantidad de dioxina necesaria para matar cada uno de n = 50 cobayas y se calcula la media de la muestra, $\bar{y}$.

(a) Calcula $\mu_{\bar{y}}$ y $\sigma_{\bar{y}}$.
(b) Determina la probabilidad de que la cantidad media de dioxina necesaria para matar los 50 cobayas sea mayor que 0.0000053 gramos.

(15º) La determinación de porcentaje de obturación de la luz solar por las copas de los árboles de un bosque es indispensable para la evaluación del hábitat de animales silvestres, la estimación de escurrimientos en las cuencas fluviales, el control de la erosión y otras actividades de silvicultura. Un método que utilizan los geocientíficos para estimar el porcentaje de obturación de las copas de los árboles se vale de un sensor de satélite llamado Landsat Thematic Mapper (trazador de mapas temáticos Landsat). Se ha realizado un estudio del porcentaje de obturación de las copas de los árboles en el Bosque Nacional San Juan (Colorado) mediante un examen de datos del simulador TMS recolectados por aviones en diversos sitios del bosque (IEEE Trans. on Geoscience and Remote Sensing, enero de 1986). Se ha calculado que la media y la desviación estándar de las lecturas obtenidas del canal 5 del TMS fueron 121.74 y 25.52, respectivamente.

(a) Sea $\bar{y}$ la lectura TMS media para una muestra de 32 sitios del bosque. Suponiendo que las cifras citadas son valores de la población, describe la distribución de muestreo $\bar{y}$.
(b) Utiliza la distribución de muestreo del apartado (a) para calcular la probabilidad de que $\bar{y}$ esté entre 118 y 130.
(c) Sea s^2 la varianza de las lecturas TMS para los 32 sitios forestales muestreados. Suponiendo que la muestra proviene de una población normal, estima la probabilidad de que s^2 sea mayor que 1311.

(16º) Una gran empresa de consultoría en arquitectura e ingeniería ha iniciado un programa de compensación de su personal gerencial por los días de incapacidad no utilizados. La empresa ha decidido pagar a cada gerente un bono por cada día de incapacidad que no aprovechó. En años anteriores, el número Y de días de incapacidad utilizados por gerente por año tuvo una distribución de probabilidad con media μ = 9.2 y varianza σ^2 = 3.24. Con el propósito de determinar si el programa de compensación redujo efectivamente el número medio de días de incapacidad no utilizados, la empresa muestreó al azar n 0 80 gerentes y registro Y (el número de días de incapacidad aprovechados por cada uno durante el año).

(a) Si se supone que el programa de compensación no redujo efectivamente el número de días de incapacidad utilizados, calcula la probabilidad de que sea menor que 8.80 días.
(b) Si se observa que $\bar{y}$ < 8.80, ¿qué inferencia harías acerca de la efectividad del programa de compensación?.

(17º) A fin de determinar si un torno para metal que produce cojinetes está debidamente ajustado, se obtiene una muestra de 36 cojinetes y se mide el diámetro de cada uno. Supón que la desviación típica del diámetro de los cojinetes medida durante un período es de 0.001 pulgadas.

(a) ¿Qué probabilidad hay de que el diámetro medio $\overline{y}$ de la muestra de 36 cojinetes esté a menos de 0.0001 pulgadas del diámetro medio de la población de cojinetes?.

(b) Supón que el diámetro medio de los cojinetes producidos por la máquina debe ser de 0.5 pulgadas. La compañía utiliza la media de la muestra para decidir si el proceso está o no en control (es decir, si está produciendo o no cojinetes con un diámetro de 0.5 pulgadas). Se considerará que la máquina está fuera de control si la media de la muestra de n = 36 diámetros es menor que 0.4994 pulgadas o mayor que 0.5006 pulgadas. Si la media verdadera del diámetro de los cojinetes producidos por la máquina es 0.501 pulgadas, ¿qué probabilidad hay de que la prueba no sugerirá que el proceso está fuera de control?.

(18º) El transporte de partículas neutrales en un ducto evacuado es un aspecto importante del diseño de reactores de fusión nuclear. En un experimento, partículas que entran por los extremos del ducto fluyeron sin impedimentos hasta chocar con la pared interior del ducto. Al chocar, fueron dispersadas (reflejadas) o bien absorbidas por la pared (Nuclear Science and Engineering, mayo de 1986). Se determinó que la probabilidad de reflexión (esto es, la probabilidad de que una partícula sea reflejada por la pared) para un tipo de ducto es de 0.16.

(a) Si dos partículas son liberadas en el ducto, calcula la probabilidad de que ambas sean reflejadas.

(b) (b) Si cinco partículas se liberan en el ducto, calcula la probabilidad de que las cinco son absorbidas.

(c) (c) ¿Qué suposición se debe hacer acerca de los sucesos simples de los apartados (a) y (b) para calcular las probabilidades?.

(d) (d) Si se liberan 2000 partículas neutras en un tipo desconocido de ducto al vacío dentro de un reactor de fusión nuclear. De éstas, 280 son reflejadas. ¿Qué probabilidad aproximada hay de que 280 o menos de las 2000 partículas neutras sean reflejadas de la pared interior del ducto si la probabilidad de reflexión del ducto al vacío es 0.16?.

(19º) Se han efectuado pruebas de cizalladura de bloque en piezas de madera estructural reparadas con resinas epóxicas indicando que la distribución de probabilidad de la resistencia de unión con grano paralelo en muestras de madera de aserrado tiene una media de 1312 libras por pulgadas cuadrada (psi) y una desviación estándar de 422 psi (Journal of Structural Engineering, febrero de 1986). Supón que se escoge al azar una muestra de 100 muestras de madera reparada con resinas epóxicas y que se mide la resistencia de unión en cada uno.

(a) Describe la distribución de muestreo $\overline{y}$, la resistencia de unión media de las muestra de 100 especímenes de madera reparada con resinas epóxicas.

(b) Calcula p($\overline{y} \geq 1418$).

(c) Si la media calculada de la muestra es en realidad $\bar{y}$=1418, ¿qué deduces acerca de los resultados de la prueba de cizallamiento de bloques?

(d) Describe la distribución muestral s^2, la varianza de las resistencias de unión de los 100 especímenes muestreados de madera reparada con resinas epóxicas. Supón que la muestra proviene de una población normal.

(e) Estima p(s > 500).

(20º) El tiempo de espera y hasta la entrega de un nuevo componente para una unidad de procesamiento de datos está **<u>distribuido uniformemente</u>** dentro del intervalo 1 a 5 días. El costo c (en cientos de dólares) de este retardo para el comprador está dado por $c = (2y^2 + 3)$. Calcula la probabilidad de que el costo de retardo sea de por lo menos 2000 dólares.

(21º) La duración Y de un componente electrónico de un microordenador doméstico tiene una densidad de Raleigh dada por

$$f(y) = \begin{cases} \dfrac{2y}{\beta} e^{-\frac{y^2}{\beta}} & si\ y > 0 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

siendo β una constante típica de cada caso. Obtén la función de densidad de probabilidad para $w = y^2$ e identifica el tipo de densidad que se obtiene con las funciones de densidad que aparecen en la bibliografía especializada.

<u>Ejercicios de repaso</u>

(22º) Los pesos de los paquetes recibidos en un departamento de almacenamiento tienen una media de 300 libras y una desviación típica de 50 libras. ¿Cuál es la probabilidad de que el peso de 25 paquetes recibidos al azar y cargados en un ascensor supere el límite de seguridad del ascensor, que es de 8200 libras?.

(23º) Ciertos tubos fabricados por una compañía tienen una duración media de 800 horas y una desviación típica de 60 horas. Halla la probabilidad de que una muestra al azar de 16 tubos, tomada de entre ellos tenga una duración media de (a) entre 790 y 810 horas, (b) menor de 785 horas, (c) mayor de 820 horas, (d) entre 770 y 830 horas.

(24º) Halla la probabilidad de que en los próximos 200 niños nacidos (a) menos del 40 % sean niños, (b) entre el 43 % y el 57 % sean niñas, (c) más del 54 % sean niños (considera para ello iguales las probabilidades de nacimiento de niño y niña). Si se consideran 100 en lugar de 200 niños, repite de nuevos todos los apartados y explica las diferencias en los resultados.

(25º) Una urna contiene 80 bolas de las que 60 % son rojas y 40 % blancas. De un total de 50 muestras de 20 bolas cada, sacadas de la urna con remplazamiento, ¿en cuántas cabe esperar (a) igual número de bolas rojas y blancas, (b) 12 bolas rojas y 8 blancas, (c) 8 rojas y 12 blancas, (d) 10 o más bolas blancas?.

(26º) A y B fabrican dos tipos de cables, que tienen unas resistencias medias de rotura de 4000 y 4500 libras con desviaciones estándar de 300 y 200 libras, respectivamente. Si se comprueban 100 cables de A y 50 cables de B, ¿cuál es la probabilidad de que la media de resistencia a la rotura de B sea (a) al menos 600 libras más que A, (b) al menos 450 libras más que A?.

(27º) El voltaje medio de una batería es de 15 voltios y la desviación típica 0,2 voltios. ¿Cuál es la probabilidad de que cuatro de estas baterías conectadas en serie tengan un voltaje conjunto e 60.8 o más voltios?.

(28º) Los resultados de unas elecciones mostraron que un cierto candidato recibió el 65 % de los votos. Determina la probabilidad de que en dos muestras al azar compuestas cada una de 200 votantes, haya una diferencia superior al 10 % en las proporciones que votaron a dicho candidato.

(29º) Una población de 7 números tiene una media de 40 y una desviación estándar de 3. Si se extraen muestras de tamaño 5 de esta población y se calcula la varianza de cada muestra, halla la media de la distribución muestral de varianzas si el muestreo es (a) con remplazamiento, (b) sin remplazamiento.

(30º) Ciertos tubos producidos por una compañía tienen una duración media de 900 horas y una desviación típica de 80 horas. La compañía despacha 100 lotes de 100 tubos cada uno. ¿En cuántos lotes cabe esperar que (a) la media de las duraciones sobrepase las 910 horas, (b) las desviación típica de las duraciones sobrepase las 95 horas?, ¿qué supuestos deben hacerse?.

Si la mediana de la duración es de 900 horas, ¿en cuántos lotes cabe esperar que la media de las duraciones sobrepase las 910 horas. Compara la respuesta con la del apartado (a) y explica los resultados.

(31º) En una ciudad las puntuaciones de un examen se distribuyen normalmente con media 72 y desviación estándar 8. (a) halla la puntuación mínima del 20 % superior de los estudiantes. (b) Halla la probabilidad de que en una muestra al azar de 199 estudiantes la puntuación mínima del 20 % superior sea menor de 76 puntos.