

# Lista de ejercicios

Curso: Introducción a la Probabilidad y Estadística CM-072

## Lecturas Importantes

1. Notas sobre Inferencia Estadística: The Element of Statistical Learning Trevor Hastie, Robert Tibshirani and Jerome Friedman.

[statweb.stanford.edu/~tibs/ElemStatLearn/](http://statweb.stanford.edu/~tibs/ElemStatLearn/).

---

1. Una población se compone de cinco números 2, 3, 6, 8, 11. Considera todas las muestras posibles de tamaño dos que pueden extraerse con reemplazamiento de esta población. Halla
  - (a) la media de la población.
  - (b) la desviación estándar de la población.
  - (c) la media de la distribución muestral de medias.
  - (d) la desviación estándar de la distribución muestral de las medias, es decir el error típico de las muestras.
2. Calcula lo mismo que en el problema anterior, pero en el caso de un muestreo sin reemplazamiento.
3. Las puntuaciones obtenidas en un test de cultura general por los alumnos de sexto curso de primaria, se distribuyen normalmente con media 100 y desviación estándar 14.2
  - (a) ¿Cuál es la probabilidad de que un alumno de sexto curso de primaria elegido aleatoriamente obtenga una puntuación mayor que 130 en el citado test?
  - (b) ¿Cuál es la probabilidad de que el alumno seleccionado obtenga una puntuación comprendida entre 90 y 115.
4. Los datos del Departamento de Agricultura indican que el consumo de manzanas de una mujer elegida aleatoriamente se distribuye según una normal de media de 19.9 libras y una desviación típica de 3.2 libras, mientras que el consumo de manzanas de un hombre elegido aleatoriamente sigue una normal con media 20.7 libras y desviación estándar 3.4 libras. Supongamos que se elige aleatoriamente a un hombre y a una mujer. ¿Cuál es la probabilidad de que el consumo de manzanas de la mujer sea mayor que el del hombre?
5. Un auditor toma una muestra aleatoria de 100 cuentas por cobrar de una población de 500 cuentas por cobrar. El auditor sabe que las 500 cuentas por cobrar constituyen una población finita cuya desviación estándar es  $\sigma = \$145$ . ¿Cuál es la probabilidad de que la media muestral difiera de la media poblacional en más de \$26?
6. Las calificaciones de un examen se distribuyen de acuerdo a una distribución normal de media 100 y desviación estándar 14.2.

¿ En qué rango de valores se encuentran el 1% más alto de las puntuaciones?
7. Un productor de tabaco mantiene que el contenido medio de nicotina de sus cigarrillos es de 2.4 miligramos con una desviación estándar 0.2 miligramos. Si se acepta que estas cifras son correctas, calcule el valor esperado y la varianza del contenido medio muestral en nicotina de 36, 64, 100, 900 cigarrillos seleccionados aleatoriamente.

8. Una compañía aseguradora de automóviles tiene 10000 aseguradores. Si el gasto anual que un asegurado le ocasiona a la compañía tiene por media 260 dólares con una desviación estándar de 800 dólares, aproxima la probabilidad de que el gasto total que la compañía debe afrontar en un año sobrepase los 2.8 millones de dólares.
9. En el problema anterior, calcula la probabilidad de que el pago anual al que debe hacer frente la compañía de seguros esté comprendido entre 2.5 y 2.7 millones de dólares.
10. Los niveles de colesterol en la sangre de una población de trabajadores tiene media 202 y desviación estándar 14.
  - (a) Si se selecciona una muestra de 36 de trabajadores, aproxime la probabilidad de que la media muestral de sus niveles de colesterol esté comprendida entre 198 y 206.
  - (b) Repite lo anterior para un tamaño muestral igual a 164.
11. Una astrónoma pretende medir, en unidades de años luz, la distancia existente entre su laboratorio y una estrella muy alejada. Sin embargo, la astrónoma sabe que, debido a las distintas condiciones atmosféricas y a los errores normales, cada vez que realiza una medida no obtiene la distancia exacta, sino sólo una estimación de ella. Como resultado, ha decidido realizar 10 mediciones distintas y utilizar la media de todas ellas como un valor estimado de la distancia real. Si los valores de las 10 mediciones realizadas constituyen una muestra extraída de la población de todas las posibles mediciones y esta población tiene media  $d$  (distancia real) y desviación estándar 3 años luz, calcule la probabilidad de que el valor estimado por la astrónoma difiera de la distancia real en menos de 0.5 años luz.
12. En el ejercicio anterior, calcula la probabilidad de que la estimación de la astrónoma difiera de la distancia real en menos de 0.5 años luz, si
  - (a) Realiza 100 observaciones.
  - (b) Lleva a cabo 10 observaciones tras descubrir un método para mejorar las mediciones, de forma que la desviación estándar de cada observación se reduzca de 3 a 2 años luz.
13. Supongamos que las baterías de un coche eléctrico funcionan durante un número de millas que tiene media  $\mu$  y desviación estándar 100. Mediante el teorema del límite central, aproxima la probabilidad de que el número medio de millas recorridas por batería, para un conjunto de  $n$  éstas difiera de  $\mu$  en más de 20 si:  $n = 10, 20, 40, 100$ .
14. Un semiconductor de circón es un elemento esencial para el funcionamiento de un superconductor y debe ser reemplazado inmediatamente cuando falle. Su duración esperada es de 100 horas y su desviación estándar es de 34 horas. Si se dispone de 22 de dichos semiconductores, calcule la probabilidad de que el semiconductor esté operativo durante las próximas 2000 horas. (Este es, calcula la probabilidad de que la suma de las duraciones de los 22 semiconductores supere las 2000 horas).
15. Supongamos que exactamente un 46% de la población está a favor de un determinado candidato. Si se extrae una muestra aleatoria de tamaño 200. ¿Cuál es la probabilidad de que al menos 100 de ellos este a favor del candidato?
16. Una compañía aérea, el porcentaje de pasajeros que tienen reserva y no se presentan es del 6%. Si existen 260 personas con reserva en un determinado vuelo que puede admitir un máximo de 250 pasajeros, aproxime la probabilidad de que la compañía sea capaz de acomodar a todos los pasajeros con reserva que aparezcan.
17. Halla la probabilidad de que en 120 lanzamientos de una moneda el número de caras
  - (a) esté comprendido entre el 40% y el 60%.
  - (b) sea  $\frac{5}{8}$  más del número de lanzamientos.
18. Cada persona de un grupo de 500 lanza una moneda 120 veces. ¿En cuántos individuos cabe esperar que

- (a) el número de caras se encuentra entre el 40% y el 60% de sus lanzamientos.
- (b)  $\frac{5}{8}$  más del número de lanzamientos resulten cara.
19. Se ha encontrado que el 2% de las piezas producidas por cierta máquina. ¿Cuál es la probabilidad de que en una partida de 400 piezas sean defectuosas
- (a) 3% o más?.
- (b) 2% o menos?.
20. Las bombillas eléctricas de un fabricante  $A$  tienen una duración media de 1,400 horas, con una desviación típica de 200 horas, mientras que la del otro fabricante  $B$  tienen una duración media de 1200 horas con una desviación estándar de 100 horas. Si se toma una muestra al azar de 125 bombillas de cada fabricante. ¿Cuál es la probabilidad de que las bombillas de  $A$  tengan una duración media que sea al menos 250 horas más que las bombillas de  $B$ ?
21. Mili y César juegan cara y sello, lanzando cada uno 50 monedas. Mili ganará el juego si consigue 5 o más caras que César, de otro modo gana César. Determina la proporción contra Mili de que gane un juego determinado.
22. La siguiente tabla muestra los porcentajes de individuos, por sexo, que tienen ciertos hábitos de conducta perjudiciales para la salud.

	Duermen 6 o menos horas al día	Fuman	Nunca desayunan	Tienen un sobrepeso superior al 30%
Hombres	22.7	32.6	25.2	12.1
Mujeres	21.4	27.8	23.6	13.7

- (a) Supongamos que se elige una muestra aleatoria de 300 hombres. Aproxima la probabilidad de que
- Al menos 75 nunca desayunen.
  - Fumen menos de 100.
- (b) Supongamos que se extrae una muestra aleatoria de 300 mujeres. Aproxima la probabilidad de que
- Al menos 25 de ellas tengan un sobrepeso superior al 30%.
  - Menos de 50 duerman 6 horas como máximo.
- (c) Supongamos que se extraen dos muestras de 300 hombres y de 300 mujeres, respectivamente. Calcula la probabilidad aproximada de que haya más fumadores en la muestra de hombres que en la de las mujeres.
- Sugerencia** Denota por  $X$  y  $Y$ , respectivamente, el número de hombres y de mujeres que fuman en cada una de las muestras. Escribe la probabilidad pedida como  $\mathbb{P}\{X - Y > 0\}$ .
23. Consideramos el siguiente modelo simple referido a los cambios diarios en el precio de una acción. Supongamos que cada día el precio de la acción pueda subir una unidad con una probabilidad de 0.52 o bajar una unidad con una probabilidad de 0.48. Si el precio de la acción en el día de hoy es de 200 y se denota  $X$  el precio que tendrá al cabo de 100 días,

- (a) Defina las variables  $X_1, X_2, \dots, X_{100}$ , tales que

$$X = 200 + \sum_{i=1}^{100} X_i$$

- (b) Obtenga  $\mathbb{E}[X_i]$ .
- (c) Calcula el valor de  $\mathbb{V}(X_i)$ .
- (d) Utilice el teorema central del límite para calcular  $\mathbb{P}\{X \geq 0\}$ .

24. Una empresa estima que el 4% el porcentaje de clientes cuyos créditos han pasado a cobranza dudosa. Un auditor revisa periódicamente las cuentas por cobrar para efectivizar la cobranza.
- Calcule aproximadamente la probabilidad de que en una muestra aleatoria de 150 clientes con cuentas por cobrar, el 6% haya pasado a cobranza dudosa.
  - El auditor acepta que el 4% el porcentaje de clientes cuyo créditos han pasado a cobranza dudosa, si cada vez que se revisa una muestra de 100 cuentas por cobrar encuentra menos de 5 cuentas que han pasado a cobranza dudosa. ¿Cuál es la probabilidad de que el auditor acepte que es 4% el porcentaje de clientes cuyos créditos han pasado a cobranza dudosa si realmente el porcentaje es 6%?
25. Los siguientes datos representan el número de cuentas registrados durante 10 días

43, 37, 50, 51, 58, 105, 52, 45, 45, 10

- Calcula la media, mediana, quartiles y desviación estándar.
  - Verifica si existe outliers usando la regla  $1.5(IQR)$ .
  - Elimina los outliers si es que los hay y calcula la media, mediana, quartiles y desviación estándar otra vez.
  - Realiza una conclusión acerca de los outliers en estadística descriptiva.
26. Un proveedor de la red investiga la carga de su red. El número de usuarios al mismo tiempo se registra en cincuenta lugares.

17.2	22.1	18.5	17.2	18.6	14.8	21.7	15.8	16.3	22.8
24.1	13.3	16.2	17.5	19.0	23.9	14.8	22.2	21.7	20.7
13.5	15.8	13.1	16.1	21.9	23.9	19.3	12.0	19.9	19.4
15.4	16.7	19.5	16.2	16.9	17.1	20.2	13.4	19.8	17.7
19.7	18.7	17.6	15.9	15.2	17.1	15.0	18.8	21.6	11.9

- Calcula la media muestral, varianza y desviación estándar del número de usuarios.
  - Estima el error estándar de la media muestral.
  - Calcula el rango intercuartil. Existe algún outliers?
  - Es reportado que el número de usuarios sigue aproximadamente una distribución normal. Apoya un histograma esta afirmación?
27. Para evaluar la efectividad de un procesador, para un cierto tipo de tareas, recogemos el tiempo del CPU de manera aleatoria para determinados trabajos (en segundos) con  $n = 30$ :

70	36	43	69	82	48	34	62	35	15
59	139	46	37	42	30	55	56	36	82
38	89	54	25	35	24	22	9	56	19

Para estos datos tenemos que si la media muestral  $\bar{X}$  y la desviación estándar  $s$ , prueba usando la desigualdad de Chebishev que al menos 8/9 de todas las tareas requiere menos que

$$\bar{X} + 3s = 127.78\text{sec}$$

de tiempo del CPU.

28. Estima la el parámetro  $\theta$  desde una muestra

3,3,3,3,7,7,7

desde una distribución discreta con un pmf

$$\begin{cases} \mathbb{P}(3) = \theta \\ \mathbb{P}(7) = 1 - \theta \end{cases}$$

Calcula dos estimadores de  $\theta$

- (a) Por el método de estimadores de momentos.
- (b) Desde el estimador de máxima probabilidad.
- (c) Estima el error estándar para cada estimador.