1. Preguntas de razonamiento (3 puntos, 35 minutos)

1.1. (1.5 puntos) *(15 minutos)* Razona detalladamente cuál sería la distribución en el muestreo del estadístico

$$T = \frac{\sum_{i=1}^{N} X_i}{N}.$$

Ten en cuenta que la población de la que procede la muestra tiene media poblacional μ y varianza poblacional σ^2, pero que no es normal. Haz todas las asunciones que creas necesarias para obtener la distribución de T. No olvides detallar los cálculos de la media y la varianza del estadístico.

1.2. (1.5 puntos) *(20 minutos)* Si X es una variable aleatoria con distribución

$$P(X = x) = p(1-p)^x \qquad \text{para } x = 0, 1, 2, \ldots,$$

razona cuál es la distribución para la variable aleatoria

$$Y = (X-2)^2.$$

2. Problemas (7 puntos, 110 minutos)

2.1. (1 punto) *(10 minutos)*

(a) En un laboratorio se analiza a un grupo de personas de las cuales se sabe que el 5 % de las cuales tienen la enfermedad de Hansen. La prueba detecta la enfermedad de Hansen al 98 % de las personas que padecen la enfermedad, pero da falsos positivos al 3 % de las que no la padecen. ¿Cuál es la probabilidad de que alguien que dé positivo en la prueba de la enfermedad de Hansen la tenga realmente?

(b) En este mismo laboratorio, se ha hecho un estudio sobre la prevalencia de otra enfermedad en cierta ciudad. De los 600 participantes en el estudio, 200 tenían la enfermedad ¿Cuál es el intervalo de confianza de la proporción de enfermos en dicha ciudad? Usa la aproximación del teorema central del límite y un nivel de confianza del 99 %. En los anexos encontrarás algunos valores útiles para resolver el ejercicio.

2.2. (1 punto) *(15 minutos)* Una casa farmacéutica es capaz de producir una determinada pastilla de una forma muy precisa. Esto es, el peso de las pastillas producidas tiene una desviación estándar muy pequeña. El equipo de investigación de la empresa ha propuesto un nuevo método de producción del fármaco. Sin embargo, esto conlleva algunos costes y se adoptará solo si hay pruebas sólidas de que la desviación estándar del peso de los artículos caerá por debajo de 0.4 miligramos (mg) con el nuevo método (¡mejor que la precisión actual!). Se produce una muestra de 10 medicamentos con el nuevo método obteniéndose que los medicamentos tienen una media de 5.724 mg y una desviación estándar de 0.35 mg. ¿Hay evidencia que apoye la adopción del nuevo método?

(a) (0.5 puntos) Indica claramente cuáles serían las hipótesis nula y alternativa, así como las asunciones que consideres necesarias para realizar este estudio.

(b) (0.5 puntos) Con un nivel de significación del 1 %, indica claramente cuál sería tu conclusión (¿debe adoptarse o no?) empleando el método de la región crítica. En el Anexo encontrarás valores útiles para realizar el ejercicio.

2.3. (1.5 puntos) *(25 minutos)* Un historiador está investigando sobre la intoxicación por plomo a lo largo de la historia. El resumen de datos mostrado en la tabla 1 se obtuvo a partir de una comparación del contenido de plomo del cabello humano extraído de individuos adultos muertos entre 1880 y 1920 con el contenido de plomo de los adultos actuales.

	Cantidad de plomo (μg) entre 1880-1920	Cantidad de plomo (μg) hoy en día
Número de muestras	30	100
Media muestral	48.5	26.6
Desviación típica muestral	14.5	12.3

Cuadro 1: Niveles de plomo a lo largo de la historia.

(a) (0.5 puntos) Al historiador le gustaría demostrar que los niveles de plomo eran mayores en la antigüedad que hoy en día. Indica claramente cuáles serían las hipótesis nula y alternativa. Indica también cuáles son las asunciones necesarias para realizar este estudio; ten en cuenta que hay evidencia que sugiere que las desviaciones estándar de ambas poblaciones son iguales.

(b) (0.375 puntos) ¿Cuál sería el p-valor del test? Haz el desarrollo matemático hasta donde sea posible y luego indica los comandos de R que te permitirían hallar el p-valor (no es necesario obtener un número).

(c) (0.25 puntos) Indica claramente (en palabras que pueda entender el historiador) cuál sería tu conclusión si el p-valor fuese 0.00012 con un nivel de significación del 1 %.

(d) (0.375 puntos) Los datos de los que se ha obtenido la tabla 1 están guardados en el *data.frame data*. Las primeras filas de este *data.frame* se muestran a continuación:

```
   plomo      periodo
1  59.92848   pasado
2  74.46782   pasado
3  40.12556   hoy
4  26.32784   hoy
```

Indica qué comandos de R te permitirían realizar el test estadístico pedido a partir del *data.frame data* (no está permitido usar las funciones estadísticas dxxx, qxxx o pxxx o rxxx).

2.4. (1.5 puntos) *(30 minutos)* Gertrude Cox va todos los días al banco. En el banco hay dos colas, una cola estándar y una cola prioritaria. Independientemente del número de personas en el banco, la cola estándar tiene un tiempo medio de espera de 10 minutos, mientras que la prioritaria tiene un tiempo medio de espera de 5 minutos. Desgraciadamente, el banco está intentando ahorrar costes y solo abre ocasionalmente la cola prioritaria. Cada día se decide si se abre la cola prioritaria con independencia de lo que haya sucedido en días anteriores. La probabilidad de abrir la cola prioritaria es 0.7. Gertrude es cliente VIP, así que tiene derecho a usar la cola prioritaria. Si Gertrude va al banco y encuentra la cola prioritaria abierta se pone en ella, en otro caso usa la cola estándar.

(a) (0.5 puntos) Sabiendo que los tiempos de espera se pueden modelar con una distribución exponencial, obtén la distribución del tiempo de espera (diario) de Gertrude.

(b) (0.5 puntos) Calcula la probabilidad de que Gertrude tenga que esperar más de 4 minutos en una visita al banco. Calcula el valor sin usar R, haciendo uso de las funciones de densidad/distribución de los chuletarios.

(c) (0.5 puntos) Calcula la probabilidad de que, en 7 visitas al banco, Gertrude sea atendida en menos de 4 minutos en tres ocasiones (Si no has logrado acabar (b), usa 0.1 como probabilidad de que Gertrude sea atendida en menos de 4 minutos).

i. Calcula el valor sin usar R, haciendo uso de las funciones de densidad/distribución de los chuletarios.

ii. Indica el (único) comando de R que tendrías que usar para calcular esta misma probabilidad.

2.5. (1 punto) **[Problema en R]** *(15 minutos)* En cierto clúster de computación se ejecutan en paralelo tres tipos de trabajos: A, B y C. Sean X_a, X_b y X_c los tiempos que transcurren hasta que cada uno de estos trabajos se terminan de ejecutar. Todos los trabajos empiezan en el mismo instante de tiempo. Por otra parte, los "tiempos de ejecución"son independientes entre sí y tienen distribuciones exponenciales con parámetros respectivos $\lambda_a = 2, \lambda_b = 3$ y $\lambda_c = 5$. Sea M el instante de tiempo en que el todos los trabajos se han completado. Escribe el código R para comprobar mediante simulaciones (no está permitido usar las funciones estadísticas dxxx, qxxx o pxxx) si la distribución de M tiene distribución exponencial de parámetro 0.95. Para ello, procede como sigue:

- Simula muchas muestras de la variable M.
- Dibuja la distribución de las muestras simuladas...
- ... y compárala con la distribución teórica (añadiendo una segunda curva al *plot* original).

2.6. (1 punto) **[Problema en R]** *(15 minutos)* Jacob Bernoulli propuso el siguiente juego de dados. El jugador paga un euro y lanza un dado. Luego lanza un conjunto de n dados, donde n es el número que muestra el primer dado. El número total de puntos que muestran los n dados se utiliza para determinar el resultado del juego. Si el número es menor que doce pierde el euro apostado, mientras que si el número es igual o mayor que doce recibe dos euros. Calcula, usando simulaciones en R, la cantidad de dinero que esperarías ganar al jugar a este juego.

Anexo: valores y prototipos de funciones útiles

qnorm(0.005)	qnorm(0.01)	qnorm(0.995)	qnorm(0.99)
-2.5758	-2.3263	2.5758	2.3263

qt(0.005,df=8)	qt(0.01,df=8)	qt(0.995,df=8)	qt(0.99,df=8)
-3.3554	-2.8965	3.3554	2.8965

qt(0.005,df=9)	qt(0.01,df=9)	qt(0.995,df=9)	qt(0.99,df=9)
-3.2498	-2.8214	3.2498	2.8214

qt(0.005,df=10)	qt(0.01,df=10)	qt(0.995,df=10)	qt(0.99,df=10)
-3.1693	-2.7638	3.1693	2.7638

qchisq(0.005,df=8)	qchisq(0.01,df=8)	qchisq(0.995,df=8)	qchisq(0.99,df=8)
1.3444	1.6465	21.955	20.0902

qchisq(0.005,df=9)	qchisq(0.01,df=9)	qchisq(0.995,df=9)	qchisq(0.99,df=9)
1.7349	2.0879	23.5894	21.666

qchisq(0.005,df=10)	qchisq(0.01,df=10)	qchisq(0.995,df=10)	qchisq(0.99,df=10)
2.1559	2.5582	25.1882	23.2093

qf(0.005,df1=8,df2=8)	qf(0.01,df1=8,df2=8)	qf(0.995,df1=8,df2=8)	qf(0.99,df1=8,df2=8)
0.1334	0.1659	7.4959	6.0289

qf(0.005,df1=9,df2=9)	qf(0.01,df1=9,df2=9)	qf(0.995,df1=9,df2=9)	qf(0.99,df1=9,df2=9)
0.1529	0.1869	6.5411	5.3511

qf(0.005,df1=10,df2=10)	qf(0.01,df1=10,df2=10)	qf(0.995,df1=10,df2=10)	qf(0.99,df1=10,df2=10)
0.171	0.2062	5.8467	4.8491