

a) Investigar (puede ser en el capítulo del Spiegel que habla sobre "Distribuciones de nombre propio") cuándo se usan las siguientes distribuciones, cómo se pueden identificar gráficamente cada una, cuál es su fórmula para la esperanza, cálculo de probabilidad y varianza, y qué factores (parámetros) se deben analizar:

1. Normal
2. Binomial
3. Geométrica
4. Hipergeométrica
5. Poisson
6. Uniforme
7. Beta
8. Exponencial
9. Bernoulli
10. Pareto

b) Resolver los siguientes ejercicios detectando con cuál distribución se resuelve:

1. En un sistema de monitoreo de tráfico en línea, se observa que, en promedio, 20 paquetes de datos son transmitidos por segundo a través de un canal particular. ¿Cuál es la probabilidad de que exactamente 25 paquetes sean transmitidos en un segundo dado? **4.46%**
2. Se tiene un sensor que mide la temperatura en un rango de 20°C a 35°C. Si la temperatura es leída al azar, ¿cuál es la probabilidad de que registre una temperatura menor a 30°C? **66.67%**
3. Un modelo de aprendizaje automático ha sido entrenado para clasificar correos electrónicos en "spam" o "no spam". Basado en pruebas previas, la precisión del modelo se modela con una distribución Beta con parámetros  $\alpha = 5$  y  $\beta = 1$ . ¿Cuál es la probabilidad de que el modelo tenga una precisión superior al 90%? **40.95%**
4. En un experimento de clasificación de imágenes, un modelo de inteligencia artificial tiene una precisión del 95% para identificar correctamente imágenes de gatos. Utilizando este modelo para clasificar una nueva imagen, ¿cuál es la probabilidad de que identifique correctamente que es un gato? **95%**
5. En un experimento de reconocimiento de voz, un modelo tiene una precisión del 80% en identificar correctamente palabras clave. Si el modelo prueba 10 palabras clave, ¿cuál es la probabilidad de que acierte exactamente 8? **30.20%**
6. El tiempo de respuesta de una consulta en una base de datos grande sigue una distribución normal, con una media de 2 segundos y una desviación estándar de 0.5 segundos. ¿Cuál es la probabilidad de que una consulta tarde más de 3 segundos? **2.28%**
7. La duración entre llegadas de mensajes a un servidor sigue una distribución exponencial con una media de 10 minutos. ¿Cuál es la probabilidad de que el siguiente mensaje llegue en menos de 5 minutos? **39.35%**

8. En un proceso de validación de transacciones, la probabilidad de que una transacción sea fraudulenta es del 0.1%. ¿Cuál es la probabilidad de que la primera transacción fraudulenta sea detectada en la décima transacción revisada? **0.099%**
9. En un lote de 100 semiconductores, 20 son defectuosos. Si se seleccionan al azar 10 semiconductores para una prueba de calidad, ¿cuál es la probabilidad de encontrar exactamente 3 defectuosos? **20.92%**
10. La distribución del ingreso de una población se modela con una distribución de Pareto, donde el parámetro de forma  $\alpha$  es 2.5 y el valor mínimo  $X_m$  es \$30,000. ¿Cuál es la probabilidad de que un individuo seleccionado al azar tenga un ingreso superior a \$100,000? **4.93%**
11. Un sistema de mensajería instantánea observa que los mensajes entre usuarios llegan con un promedio de 15 segundos entre cada uno. ¿Cuál es la probabilidad de que el tiempo hasta el próximo mensaje sea de al menos 10 segundos? **51.34%**
12. En un sistema de detección de anomalías, la probabilidad de detectar una señal anómala en cualquier revisión es del 5%. ¿Cuál es la probabilidad de que la primera detección ocurra en la octava revisión? **3.49%**
13. La distribución de ingresos de una cierta región se modela con una distribución de Pareto, con un parámetro de forma de 3 y un ingreso mínimo de \$20,000. ¿Cuál es la probabilidad de que una persona elegida al azar tenga un ingreso mayor a \$50,000? **6.4%**
14. En un sitio web de comercio electrónico, se observa que llegan en promedio 2 pedidos por minuto. ¿Cuál es la probabilidad de que lleguen 5 pedidos en el próximo minuto? **3.61%**
15. Un análisis de mercado muestra que la tasa de clics (CTR) para un nuevo anuncio en línea sigue una distribución Beta con parámetros  $\alpha = 2$  y  $\beta = 5$ . ¿Cuál es la probabilidad de que el CTR sea superior al 75%? **0.46%**

- c) Las pruebas con Z, t-Student, Chi-cuadrada y F de Fisher son herramientas específicas para diferentes situaciones y tipos de datos. Recopilar información sobre:
- a. ¿Cuándo se usan?
  - b. ¿En qué consiste cada una?
  - c. Fórmulas relacionadas con cada prueba.
- d) Buscar un ejercicio relevante para la Ciencia de Datos de cada una de las cuatro pruebas y resolverlo paso a paso.