



Trabajo Práctico Grupal: Análisis de Variables Aleatorias

Objetivo

Los alumnos trabajarán con un conjunto de datos reales (más de 200 registros) para identificar variables aleatorias discretas y continuas, realizar visualizaciones gráficas, comparar distribuciones y aplicar conceptos básicos de probabilidad. El trabajo culminará con la entrega de un informe y una exposición en mesa de examen.

Consignas

1. Selección de la Base de Datos

- Buscar una base de datos real (ej: datasets públicos como Kaggle, Google Dataset Search, o datos de sistemas institucionales).
- La base debe tener al menos **200 registros** y definir al menos **una variable discreta** y **una continua**.
- Ejemplos:
 - Tiempos de respuesta de un servidor (continua) vs. Cantidad de solicitudes por hora (discreta).
 - Precios de productos (continua) vs. Número de ventas por categoría (discreta).

2. Análisis Descriptivo

- Definir las variables aleatorias discretas y continuas seleccionadas.
- Calcular para cada variable:
 - **Discretas:** Media, varianza, moda y tabla de frecuencias.
 - **Continuas:** Media, desviación estándar, histograma con intervalos adecuados (regla de Sturges).
- Generar gráficos comparativos:
 - Diagrama de barras para discretas.
 - Histograma + curva de densidad para continuas (usar herramientas como Excel, Python o R).

3. Distribución de Probabilidad para Variables Continuas

- Para una variable continua:
 - **Aproximar la probabilidad sin integrar:** Usar el histograma para calcular frecuencias relativas como probabilidades (ej: $P(a \leq X \leq b) = \frac{\text{frecuencia en intervalo}}{n}$).
 - **Comparar con una distribución normal:** Superponer la curva $N(\mu, \sigma)$ sobre el histograma y discutir si los datos siguen esta distribución (usar herramientas como Excel/Google Sheets para generar la curva).
 - Explicar cómo se calcularía teóricamente la probabilidad (integral) y por qué la aproximación es útil en este contexto.

4. Conclusiones y Aplicación

- Relacionar los hallazgos con situaciones reales en sistemas (ej: predecir carga de servidores, optimizar recursos).
- Discutir limitaciones de los modelos usados.

Formato de Entrega

- **Informe PDF:** Máximo 10 páginas con código (si aplica), gráficos y explicaciones.
- **Exposición:** 15 minutos por grupo en mesa de examen (defensa de metodología y resultados).

Recomendaciones

- Usar herramientas accesibles: Excel, Google Sheets, o Python.
- Para variables continuas, enfocarse en la interpretación visual (histogramas) más que en cálculos complejos.

Preguntas Orientadoras

- ¿Qué aprendieron sobre el comportamiento de los datos?
- ¿Qué decisiones podrían tomarse a partir de este análisis?
- ¿Qué limitaciones encontraron en los datos?