

# Instrucciones para el Análisis de Muestras de Números

Prof. Eddy Ramírez

## 1. Análisis de Muestras Seleccionadas

Dadas las 4 muestras de números seleccionadas, realice los siguientes pasos para analizar su distribución y validar las distribuciones de probabilidad que las generaron.

### 1.1. Generación de Gráficos de Frecuencias

Para cada una de las 4 muestras:

- Calcule las frecuencias absolutas y relativas de los valores en la muestra.
- Genere un gráfico de barras o histograma que represente las frecuencias. Asegúrese de que el eje horizontal muestre los intervalos o clases de los datos, y el eje vertical las frecuencias.
- Etiquete adecuadamente los ejes y el título del gráfico, por ejemplo: “Histograma de Frecuencias para la Muestra 1”.

### 1.2. Gráficos de Distribuciones de Probabilidades

Despejando las distribuciones de probabilidad que generaron las muestras (por ejemplo, distribución uniforme, normal, poisson, binomial, geométrica etc.), proceda como sigue para cada muestra:

- Identifique la distribución de probabilidad asociada (parámetros como media  $\mu$ , desviación estándar  $\sigma$ , o rangos si aplica).
- Genere la función de densidad de probabilidad (PDF) o función de distribución acumulativa (CDF) de la distribución teórica.

- Superponga el gráfico de la distribución teórica sobre el histograma de frecuencias de la muestra correspondiente para visualizar la comparación.
- Ejemplo para una distribución normal: La PDF se representa como

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Etiquete los gráficos claramente, indicando la distribución y sus parámetros.

### 1.3. Pruebas Estadísticas de Bondad de Ajuste

Para corroborar que las muestras provienen de las distribuciones de probabilidad especificadas, realice las siguientes pruebas para cada una de las 4 muestras:

#### 1.3.1. Prueba de Chi-Cuadrado ( $\chi^2$ )

- Divida los datos en clases o intervalos de igual amplitud.
- Calcule las frecuencias observadas ( $O_i$ ) en cada clase y las frecuencias esperadas ( $E_i$ ) bajo la distribución teórica (usando la PDF o CDF para estimar probabilidades).
- Compute el estadístico de la prueba:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

donde  $k$  es el número de clases.

- Determine los grados de libertad ( $df = k - 1 - m$ , donde  $m$  es el número de parámetros estimados) y compare con el valor crítico de la tabla  $\chi^2$  o calcule el  $p$ -valor.

- Interprete: Si  $p > 0,05$  (nivel de significancia típico), no rechace la hipótesis nula (la muestra sigue la distribución).

### 1.3.2. Prueba de Kolmogorov-Smirnov (K-S)

- Calcule la función de distribución empírica (CDF empírica) de la muestra:  $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ , donde  $n$  es el tamaño de la muestra e  $I$  es la función indicadora.
- Obtenga la CDF teórica  $F(x)$  de la distribución especificada.
- Compute el estadístico K-S:  $D = \sup_x |F_n(x) - F(x)|$  (la máxima diferencia absoluta).
- Compare  $D$  con el valor crítico de la tabla K-S para el tamaño de muestra  $n$  y nivel de significancia  $\alpha = 0,05$ , o calcule el  $p$ -valor.
- Interprete: Si  $D$  es menor que el crítico o  $p > 0,05$ , no rechace la hipótesis nula.

## 1.4. Presentación

En una presentación en beamer final, resuma los resultados de los gráficos y pruebas para cada muestra. Discuta si las distribuciones teóricas se ajustan bien a los datos observados basándose en los resultados de las pruebas. Incluya todos los gráficos y tablas de resultados (por ejemplo, tabla de frecuencias observadas vs. esperadas para  $\chi^2$ ).

Debe encontrar la distribución que mejor se ajuste para cada caso.

## 2. Evaluación

La evaluación está dada por el encontrar la distribución que generaron cada una de las muestras  $M_i$ . Para ello debe asegurarse que pase las pruebas  $\chi^2$  y KS con la *mayor* holgura posible.

Si dos grupos tienen la misma muestra y distribuciones diferentes, los puntos se lo llevan quienes tengan mejor pruebas de bondad y ajuste.

Muestra	Valor
$M_i$	30 %
$M_i$	30 %
$M_{15}$	20 %
$M_{16}$	20 %

Toda la información evaluable debe estar presente en la presentación de beamer, o no será tomada en cuenta en la evaluación.