

## Enunciado del Proyecto Práctico PyE Sem.2-2024 (Análisis de Datos)

Usted dispone de dos muestras aleatorias (individuales y personalizadas) con 8.000 observaciones c/u de las v.a.  $W$  (discreta) ,  $(X,Y)$  (conjuntas continuas) en sendos archivos de datos llamados “MuestraW#.csv” , “MuestraXY#.csv” respectivamente, siendo # un consecutivo desde 1 hasta la cantidad de estudiantes inscritos en la asignatura según listado CONEST.

### Para la muestra dada de la v.a. $W$ :

- Calcule las medidas descriptivas muestrales : Mínimo, Máximo, Promedio, Moda, Varianza  $S^2$  de los valores observados de la v.a.  $W$
- Considerando las distribuciones :  $\text{Bin}(n;q)$  ,  $\text{DUnif}(a;b)$  ,  $\text{Geom}(q)$  ,  $\text{NegBin}(n;q)$  ,  $\text{Poisson}(\lambda)$  proponga cuál de dichas distribuciones de probabilidad es la que mejor se ajusta a la muestra dada de la v.a.  $W$  , escriba su *fdp* e indique valor(es) de su(s) parámetro(s) eligiendo entre los posibles valores : 0, 1/9, 1/8, 1/7, 1/6, 1/5, 2/9, 1/4, 2/7, 1/3, 3/8, 2/5, 3/7, 4/9, 1/2, 5/9, 4/7, 3/5, 5/8, 2/3, 5/7, 3/4, 7/9, 4/5, 5/6, 6/7, 7/8, 8/9, 1, 2, 3, 4, 5, 6, 7, 8, 9 (no vale usar otro valores)
- Dibuje un gráfico de barras comparativo exacto (sólo uno) con las frecuencias relativas observadas en la muestra dada junto con la correspondiente *fdp* de la distribución hipotética (teórica) propuesta por Usted para la v.a.  $W$  e indique sus medidas descriptivas poblacionales : Mínimo, Máximo, Media  $E(W)$ , Moda, Varianza  $\text{Var}(W)$  en un cuadro comparativo, junto con las medidas muestrales.

### Para la muestra dada de v.a. conjuntas $(X,Y)$ :

- Calcule las medidas descriptivas muestrales : Mínimo, Máximo, Promedio, Mediana, Varianza  $S^2$  de los valores observados de la v.a.  $X$
- Dibuje el histograma de frecuencias relativas observadas en la muestra de la v.a.  $X$  , agrupando los datos adecuadamente
- Considerando las distribuciones :  $\text{Exp}(\lambda)$  ,  $\text{Gamma}(k;\lambda)$  ,  $N(\mu;\sigma^2)$  ,  $\text{Triang}(a;c;b)$  ,  $\text{Unif}(a;b)$  , proponga cuál de dichas distribuciones de probabilidad es la que mejor se ajusta a la muestra dada de la v.a.  $X$  escriba su *fdp* e indique valor(es) de su(s) parámetro(s) eligiendo entre los posibles valores : 0, 1/9, 1/8, 1/7, 1/6, 1/5, 2/9, 1/4, 2/7, 1/3, 3/8, 2/5, 3/7, 4/9, 1/2, 5/9, 4/7, 3/5, 5/8, 2/3, 5/7, 3/4, 7/9, 4/5, 5/6, 6/7, 7/8, 8/9, 1, 2, 3, 4, 5, 6, 7, 8, 9 (no vale usar otros valores)
- Dibuje la gráfica exacta de la correspondiente *fdp* de la distribución hipotética (teórica) propuesta por Usted para la v.a.  $X$  e indique sus medidas descriptivas poblacionales : Mínimo, Máximo, Media  $E(X)$ , Mediana, Varianza  $\text{Var}(X)$  en un cuadro comparativo, junto con las medidas muestrales
- Dibuje un gráfico de dispersión (nube de puntos) con los valores observados de las v.a.  $(X,Y)$  calcule su Coeficiente de Correlación Muestral  $r$ , estime además los coeficientes  $a,b$  de la correspondiente Recta de Regresión  $E(Y|X)=aX+b$ . En este caso particular: Es útil conocer el valor de  $X$  para estimar  $E(Y|X)$  ? Justifique formalmente y explique brevemente vuestra respuesta

Para este trabajo puede usar RStudio y/o una hoja de cálculo, en cuyo caso debe anexar archivo(s) con vuestro nombre “ProyectoAnaGil.r” y/o “ProyectoAnaGil.xlsx” además de “ProyectoAnaGil.pdf”

Este trabajo es estrictamente individual, así que la falta de originalidad será penalizada severamente