

### Tarea 3 de Probabilidad y Estadística aplicada

El objetivo de esta tarea es simular las variables aleatorias discretas vistas en el curso usando Python para obtener muestras con las que hacer estadística descriptiva y una verificación empírica de la ley de los grandes números.

Trabajaremos con tres distribuciones discretas:

- Distribución 1: binomial de parámetros  $n = 100$ ,  $p = 0,35$ . Se puede usar el submódulo `binom` de la librería `scipy.stats` para generar una muestra aleatoria simple con esta distribución, por ejemplo

```
from scipy.stats import binom

r = binom.rvs(n,p,size = 1000)
```

- Distribución 2: geométrica de parámetro  $p = 0,08$ . Se puede usar el submódulo `geom` de la librería `scipy.stats` para generar una muestra aleatoria simple con esta distribución, por ejemplo

```
from scipy.stats import geom

r = geom.rvs(p,size = 1000)
```

- Distribución 3: poisson de parámetro  $\lambda = 30$ . Se puede usar el submódulo `poisson` de la librería `scipy.stats` para generar una muestra aleatoria simple con esta distribución, por ejemplo

```
from scipy.stats import poisson

r = poisson.rvs(L,size = 1000)
```

Se pide:

**\*\*1 media empirica y med. teorica = lim tiende a la varianza teorica.**

1. En este ejercicio trabajaremos con la distribución 1.

- a) Generar muestras aleatorias de tamaños  $10^2$ ,  $10^3$ ,  $10^4$  y  $10^5$ .
- b) Hacer un diagrama de cajas para cada una de las muestras generadas en la parte anterior. ¿Existen datos atípicos en las muestras?
- c) Realizar un histograma de las muestras generadas.
- d) Hallar la mediana y la moda de cada muestra. **estadística descriptiva**
- e) Hallar la media empírica de cada muestra y compararla con la esperanza teórica de la distribución 1. ¿Qué se puede observar en las muestras más grandes? **\*\*1**
- f) Hallar la varianza empírica de cada muestra y compararla con la varianza teórica de la distribución 1. ¿Qué se puede observar en las muestras más grandes?

que cuanto más grande el numero, más se acerca a la teorica

2. En este ejercicio trabajaremos con la distribución 2.

- a) Generar muestras aleatorias de tamaños  $10^2$ ,  $10^3$ ,  $10^4$  y  $10^5$ .
- b) Hacer un diagrama de cajas para cada una de las muestras generadas en la parte anterior. ¿Existen datos atípicos en las muestras?

- c) Realizar un histograma de las muestras generadas.
  - d) Hallar la mediana y la moda de cada muestra. **estadística descriptiva**
  - e) Hallar la media empírica de cada muestra y compararla con la esperanza teórica de la distribución 2. ¿Qué se puede observar en las muestras más grandes?
  - f) Hallar la varianza empírica de cada muestra y compararla con la varianza teórica de la distribución 2. ¿Qué se puede observar en las muestras más grandes?
3. En este ejercicio trabajaremos con la distribución 3.
- a) Generar muestras aleatorias de tamaños  $10^2$ ,  $10^3$ ,  $10^4$  y  $10^5$ .
  - b) Hacer un diagrama de cajas para cada una de las muestras generadas en la parte anterior. ¿Existen datos atípicos en las muestras?
  - c) Realizar un histograma de las muestras generadas.
  - d) Hallar la mediana y la moda de cada muestra.
  - e) Hallar la media empírica de cada muestra y compararla con la esperanza teórica de la distribución 3. ¿Qué se puede observar en las muestras más grandes?
  - f) Hallar la varianza empírica de cada muestra y compararla con la varianza teórica de la distribución 3. ¿Qué se puede observar en las muestras más grandes?

### Sobre el informe:

- El tiempo para entregar el informe es hasta el sábado 3 de junio inclusive. La entrega se realizará por webassignatura.
- El informe deberá estar en formato pdf, la entrega también deberá incluir los scripts utilizados.
- El informe deberá contener título, fecha, nombre y cédula de los estudiantes.
- Se evaluará: prolijidad del informe, utilización correcta del idioma español, redacción, prolijidad del código presentado en los scripts, conclusiones.

### RESPUESTAS:

Cuando las muestras son más grandes, es de esperar que los resultados de comparar la media empírica y la media teórica se aproximen más entre sí. Esto se debe a la Ley de los Grandes Números, que establece que a medida que el tamaño de la muestra aumenta, la media empírica tiende a converger hacia la media teórica.

En general, cuando las muestras son más grandes:

1- La media empírica tiende a ser más precisa y se acerca más a la media teórica.

2- La varianza de la media empírica disminuye, lo que significa que los valores se agrupan más alrededor de la media teórica.

Esto se debe a que a medida que el tamaño de la muestra aumenta, se reducen los efectos de la variabilidad aleatoria en los datos y la estimación de la media se vuelve más precisa.

Es importante destacar que, aunque se espera una mayor cercanía entre la media empírica y la media teórica con muestras más grandes, esto no garantiza una igualdad perfecta. Siempre existe la posibilidad de que haya diferencias debido a factores como el sesgo muestral, errores de medición o distribuciones no ideales.

En resumen, al aumentar el tamaño de la muestra, se espera una mayor precisión y convergencia entre la media empírica y la media teórica, lo que proporciona una mejor estimación de los parámetros poblacionales.