

UNIDAD 4

ACTIVIDAD PRÁCTICA

DETECTIVE DE DATOS:

BINOMIAL VS POISSON

ÍNDICE

| | |
|------------------------------------------------------|---|
| ORGANIZACIÓN DE LA ACTIVIDAD PRÁCTICA | 3 |
| DESCRIPCIÓN DE LA ACTIVIDAD PRÁCTICA..... | 3 |
| Identificación del origen de las muestras..... | 3 |
| Aproximación Normal de la distribución Binomial..... | 4 |
| Video-Presentación | 4 |
| RÚBRICA DE CORRECCIÓN | 5 |
| CÓMO REALIZAR LA ENTREGA | 6 |

ORGANIZACIÓN DE LA ACTIVIDAD PRÁCTICA

| Detective De Datos: Binomial Vs Poisson | |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| Tipo de tarea | Individual |
| Entregables | Un archivo INSD_PRES_U4_Nombre_Apellido1_Apellido2.zip que contenga: - detective_datos.R - detective_datos.mp4 |

DESCRIPCIÓN DE LA ACTIVIDAD PRÁCTICA

Importante: Si se utilizan argumentos o código no basados en el material proporcionado en la asignatura, se deberá indicar obligatoriamente las referencias bibliográficas (libros, páginas web, videos) utilizadas. En caso contrario, la calificación de la actividad será de 0 puntos.

El objetivo de esta actividad es analizar dos muestras de datos de origen desconocido, identificar el modelo probabilístico que mejor describe cada una de ellas y comprobar hasta qué punto podemos aproximar una distribución binomial mediante una distribución normal.

Se te proporciona el archivo `muestras.RData`, que contiene dos muestras aleatorias:

- `muestra_X`
- `muestra_Y`

Cada una de estas muestras procede de una distribución distinta:

- Distribución **Binomial($n = 12, p = 0.5$)**
- Distribución **Poisson($\lambda = 6$)**

Para cargar los datos, utiliza el siguiente bloque de código al inicio de tu script `detective_datos.R`:

```
# Carga de datos
load("muestras.RData")
# A partir de aquí, muestra_X y muestra_Y están disponibles
```

Identificación del origen de las muestras

Tu primer objetivo es decidir cuál de las dos muestras (`muestra_X` o `muestra_Y`) proviene de la distribución **Binomial($n = 12, p = 0.5$)** y cuál de la distribución **Poisson($\lambda = 6$)**.

Para ello deberás obtener **medidas estadísticas y gráficos estadísticos** de cada muestra y comparar los resultados con las propiedades teóricas de las distribuciones **Binomial($n = 12, p = 0.5$)** y **Poisson($\lambda = 6$)**.

Se recomienda que, al menos, tengas en cuenta:

- La relación entre las medidas estadísticas de las muestras y las medidas características propias de cada distribución.
- La relación entre la función de masa de probabilidad de cada distribución teórica y la frecuencia (relativa) de cada valor observado en las muestras.

Aproximación Normal de la distribución Binomial

Una vez identificada la muestra que consideras que procede de la distribución **Binomial($n = 12, p = 0.5$)**, estudiarás hasta qué punto puede aproximarse mediante una distribución Normal.

Para ello deberás:

1. Comprobar si se dan las condiciones para que la aproximación Normal sea adecuada.
2. Determinar los parámetros de la aproximación Normal (media μ y varianza σ^2) para la distribución **Binomial($n = 12, p = 0.5$)** considerada.
3. Construir una visualización que permita comparar:
 - La distribución de la muestra binomial.
 - La curva de densidad de la distribución Normal elegida como aproximación.

Video-Presentación

En la video-presentación deberás mostrar:

- Tu imagen (webcam) y tu audio durante toda la exposición.
- Dos o más diapositivas que recojan de forma clara:
 - las evidencias (medidas estadísticas, gráficos estadísticos y propiedades teóricas) utilizadas para identificar el origen de las muestras,
 - y la visualización que te ha permitido comparar la aproximación Normal y la muestra binomial.

Tu exposición oral se dividirá en dos partes:

1. **Identificación de las muestras:** Explica de forma justificada qué muestra has identificado como procedente de la distribución **Binomial($n = 12, p = 0.5$)** y cuál como procedente de la distribución **Poisson($\lambda = 6$)**.
2. **Aproximación Normal:** Comenta si consideras que la **aproximación Normal** describe razonablemente la forma de la distribución observada. ¿Cuál es la principal diferencia entre utilizar la distribución **Binomial($n = 12, p = 0.5$)** o la **aproximación Normal** para describir un fenómeno cuya distribución de probabilidad tenga esa forma?

La video-presentación debe tener una **duración total inferior a 4 minutos** y deberá entregarse en un fichero denominado **detective_datos.mp4**.

Importante: No redactes ni leas un guion de forma literal. La finalidad del video es valorar tu comprensión del análisis realizado y tu capacidad para explicar e interpretar los resultados con tus propias palabras.

RÚBRICA DE CORRECCIÓN

| Criterios | Excelente | Satisfactorio | No satisfactorio | Insuficiente |
|-----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Identificación distribuciones (6 puntos) | 4.5 – 6.0 puntos Las dos distribuciones se identifican correctamente y el análisis es riguroso, utilizando de manera adecuada las medidas estadísticas, los gráficos y las propiedades teóricas. La justificación es clara y detallada, integrando bien el trabajo en R y la explicación del vídeo, con un discurso fluido y sin lectura literal de un guion. | 3.0 – 4.4 puntos Las dos distribuciones se identifican correctamente, pero la justificación presenta alguna imprecisión o falta de profundidad. Se usan medidas y gráficos adecuados, aunque la conexión con la teoría podría estar mejor desarrollada. La explicación en el vídeo es clara, pero algo genérica o con cierta dependencia de un guion. | 1.5 – 2.9 puntos Las distribuciones están mal identificadas o la conclusión es poco clara. Se aprecia un intento de análisis con algunas medidas y gráficos, pero el uso de la teoría es limitado o contiene errores relevantes. La explicación en el vídeo es superficial o no conecta bien con los resultados obtenidos. | 0 – 1.4 puntos No se identifican las distribuciones o la respuesta es incoherente. La justificación está ausente o muestra un claro desconocimiento de los conceptos. La video-presentación no permite valorar la comprensión porque no trata este aspecto o lo hace de forma claramente insuficiente. |
| Aproximación normal: condiciones y parámetros (2 puntos) | 1.5 – 2.0 puntos Se comprueban de forma explícita y correcta las condiciones necesarias para usar la aproximación normal. Los parámetros de la distribución normal aproximante se calculan sin errores y el código en R es funcional y está razonablemente bien documentado. | 1.0 – 1.4 puntos Se comprueban las condiciones de forma básicamente correcta, aunque pueda faltar algún comentario o detalle. Los parámetros de la distribución normal se calculan de forma correcta salvo errores menores y el código funciona, aunque podría estar mejor organizado o comentado. | 0.5 – 0.9 puntos Las condiciones se mencionan de manera incompleta o con errores importantes. El cálculo de los parámetros de la distribución normal es parcial o presenta fallos significativos y el código en R tiene errores o está incompleto. | 0.0 - 0.4 puntos No se comprueban las condiciones de aproximación o se hace de forma claramente incorrecta. No se calculan los parámetros de la distribución normal o los valores son incompatibles con la distribución estudiada; el código no es funcional o no se incluye. |

| Aproximación normal: visualización y comentario (2 puntos) | 1.5 – 2.0 puntos | 1.0 – 1.4 puntos | 0.5 – 0.9 puntos | 0.0 - 0.4 puntos |
|------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| | Genera una visualización correcta que permite comparar las distribuciones. En el video, comunica con claridad sus hallazgos y ofrece una reflexión crítica sobre la calidad del ajuste observado. | La visualización es adecuada, pero presenta aspectos formales mejorables. La exposición en video describe los resultados correctamente, pero la reflexión sobre el ajuste es superficial o meramente descriptiva. | El gráfico presenta errores conceptuales que dificultan la comparación. La exposición en video es confusa, leída literalmente o incompleta. | No se entrega el video, el gráfico no se ha generado correctamente o la calidad del material entregado impide evaluar la comprensión del alumno. |

CÓMO REALIZAR LA ENTREGA

Deberás entregar un archivo `INSD_PRES_U4_Nombre_Apellido1_Apellido2.zip` que contenga:

1. El fichero `detective_datos.R` con tu código completo y comentado.
2. El video `detective_datos.mp4` con una duración inferior a 4 minutos.

Nota: Para realizar la video-presentación puedes utilizar Microsoft Teams o cualquier otro software apropiado. Si el vídeo es demasiado pesado para subirlo directamente a Blackboard, súbelo al OneDrive de vuestra cuenta (@live.u-tad.com) y enlázalo en la entrega de Blackboard.