Exámen Final SOL114

Instrucciones: Responde a las preguntas proporcionadas, asegurándote de justificar tus respuestas con cálculos y razonamientos claros.

I. En el Antiguo Testamento, el libro del Génesis comienza así: *"Y creó Dios al hombre a su imagen, a imagen de Dios lo creó; varón y hembra los creó"* (Génesis 1:27).

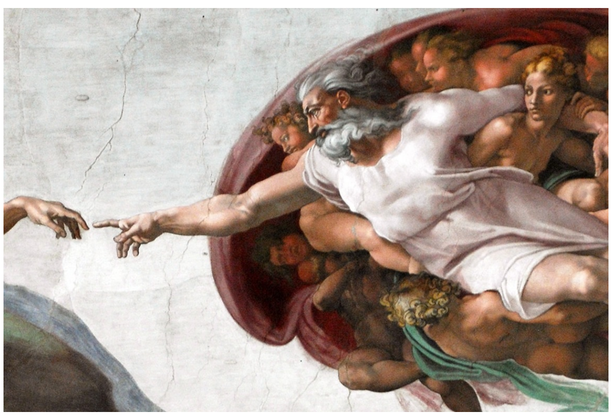


Figure 1: La creación

En esta pregunta, reemplazaremos la idea de creación divina por una versión probabilística: **variables aleatorias**.

I.i. Sea X una variable aleatoria que describe el sexo de una persona, donde la probabilidad de ser "varón" ($X = 1$) es p, y la probabilidad de ser "hembra" ($X = 0$) es $1-p$. Identifica qué distribución sigue la variable aleatoria Y y escribe su función de masa de probabilidad (PMF).

Respuesta:

La variable aleatoria X, que describe el sexo de una persona, donde $P(X = 1) = p$ (varón) y $P(X = 0) = 1 - p$ (hembra), sigue una **distribución de Bernoulli**.

La función de masa de probabilidad (PMF, por sus siglas en inglés) de una variable de Bernoulli está dada por:

$$f_X(x) = P(X = x) = \begin{cases} p & \text{si } x = 1, \\ 1 - p & \text{si } x = 0, \\ 0 & \text{en otro caso.} \end{cases}$$

En términos generales, la PMF se puede escribir como:

$$f_X(x) = p^x (1 - p)^{1-x}, \quad x \in \{0, 1\}.$$

I.ii. Si la probabilidad de ser "varón" y "hembra" es la misma, ¿cuál es el valor de p?

Respuesta:

Si la probabilidad de ser "varón" ($X = 1$) y de ser "hembra" ($X = 0$) es la misma, tenemos:

$$P(X = 1) = P(X = 0) \implies p = 1 - p.$$

Resolviendo esta ecuación:

$$p + p = 1 \implies 2p = 1 \implies p = \frac{1}{2}.$$

Por lo tanto, cuando la probabilidad es la misma, el valor de p es:

$$p = \frac{1}{2}.$$

I.iii. Si Adán (X_1) y Eva (X_2) son dos realizaciones independientes de X, expresa y calcula la probabilidad de generar dos personas secuencialmente y obtener un "varón" y una "hembra" (en ese orden).

Respuesta:

Sean X_1 y X_2 las variables independientes que representan las realizaciones de Adán y Eva, respectivamente. Deseamos calcular la probabilidad de obtener un "varón" ($X_1 = 1$) seguido por una "hembra" ($X_2 = 0$):

$$P(X_1 = 1 \text{ y } X_2 = 0).$$

Dado que X_1 y X_2 son independientes, la probabilidad conjunta se puede expresar como el producto de las probabilidades marginales:

$$P(X_1 = 1 \text{ y } X_2 = 0) = P(X_1 = 1) \cdot P(X_2 = 0).$$

Sustituyendo las probabilidades, tenemos:

$$P(X_1 = 1) = p, \quad P(X_2 = 0) = 1 - p.$$

Por lo tanto:

$$P(X_1 = 1 \text{ y } X_2 = 0) = p \cdot (1 - p).$$

Si $p = \frac{1}{2}$ (como calculamos en la parte anterior), entonces:

$$P(X_1 = 1 \text{ y } X_2 = 0) = \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

II. Asimismo, el creador probabilístico estableció que la variable aleatoria Y describe la altura (en cms) de un "varón" y de una "hembra" de la siguiente manera:

- $Y \mid X = 1 \sim \text{Normal}(\mu_v = 170, \sigma_v = 20)$
- $Y \mid X = 0 \sim \text{Normal}(\mu_h = 160, \sigma_h = 10)$

II.i. ¿Son sexo y altura dos variables independientes? Fundamenta tu respuesta.

La altura Y sigue una distribución condicional dependiente del sexo X. Esto implica que Y no tiene la misma distribución para todos los valores de X. En otras palabras, la probabilidad de Y depende del valor de X.

Por lo tanto, **sexo y altura no son independientes**, ya que la altura depende del sexo.

II.ii. Calcula la probabilidad de que un hombre tenga una altura superior a 180 cms.

La altura de un hombre ($Y \mid X = 1$) sigue una distribución normal:

$$Y \mid X = 1 \sim \mathcal{N}(\mu_v = 170, \sigma_v = 20).$$

Queremos calcular:

$$P(Y > 180 \mid X = 1).$$

Estandarizamos Y usando la transformación $Z = \frac{Y-\mu_v}{\sigma_v}$, donde $Z \sim \mathcal{N}(0,1)$. Así, la probabilidad se convierte en:

$$P(Y > 180 \mid X = 1) = P\left(Z > \frac{180 - 170}{20}\right) = P(Z > 0.5).$$

La probabilidad acumulada de $Z \leq 0.5$ se obtiene de la tabla de la distribución normal estándar:

$$P(Z \leq 0.5) = 0.6915.$$

Por lo tanto:

$$P(Z > 0.5) = 1 - P(Z \leq 0.5) = 1 - 0.6915 = 0.3085.$$

La probabilidad de que un hombre tenga una altura superior a 180 cm es:

$$P(Y > 180 \mid X = 1) \approx 0.3085.$$

II.iii. ¿Que altura debe tener una mujer para ubicarse en el percentil 90 (10% más alto)?

Para una mujer ($Y \mid X = 0$), la altura sigue una distribución normal:

$$Y \mid X = 0 \sim \mathcal{N}(\mu_h = 160, \sigma_h = 10).$$

Queremos encontrar y tal que:

$$P(Y \leq y \mid X = 0) = 0.9.$$

El percentil 90 de una distribución normal estándar (Z) corresponde a un valor acumulado de 0.9. De las tablas de la normal estándar:

$$P(Z \leq 1.28) = 0.9.$$

Desestandarizamos Z para encontrar y:

$$y = \mu_h + Z \cdot \sigma_h = 160 + 1.28 \cdot 10 = 160 + 12.8 = 172.8.$$

Por lo tanto, una mujer debe tener una altura de:

$$y = 172.8\,\text{cm},$$

para ubicarse en el percentil 90 (10% más alto).

III. Teomémas Asintóticos

III .i. Identifica los siguientes teorémas y explica en palabras haciendo referencia a los términos usados en las ecuaciones respectivas.

- a) $\bar{X} \xrightarrow{d} \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Este enunciado corresponde al **Teorema Central del Límite (TCL)**. Este teorema establece que, para una muestra aleatoria de tamaño n, si las observaciones son independientes y provienen de una población con media finita μ y varianza finita σ^2, entonces la media muestral $\bar{X}$ sigue una distribución aproximadamente normal cuando n es grande, incluso si los datos originales no tienen una distribución normal.

- $\bar{X}$: Representa la media muestral.
- $\xrightarrow{d}$: Denota convergencia en distribución (la distribución de $\bar{X}$ se aproxima a una distribución normal a medida que $n \to \infty$).
- μ: Media poblacional.
- $\frac{\sigma}{\sqrt{n}}$: Desviación estándar de la media muestral (conocida como error estándar de la media).

En palabras simples, el TCL nos permite realizar inferencias sobre la media poblacional usando la distribución normal estándar, independientemente de la distribución original de los datos, siempre que el tamaño muestral sea suficientemente grande.

- b) $\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X} - \mu\right| \geq \epsilon\right) = 0$

Este enunciado corresponde a la **Ley de los Grandes Números (LGN)**, específicamente la **versión débil de la LGN**. Este teorema garantiza que, a medida que el tamaño de la muestra n aumenta indefinidamente, la media muestral $\bar{X}$ converge en probabilidad a la media poblacional μ.

- $\mathbb{P}$: Denota probabilidad.
- $\left|\bar{X} - \mu\right| \geq \epsilon$: Evento en el cual la media muestral $\bar{X}$ difiere de la media poblacional μ por una cantidad mayor o igual a un umbral $\epsilon > 0$.
- $\lim_{n\to\infty}$: Nos indica el comportamiento cuando el tamaño muestral crece indefinidamente.

En palabras simples, la LGN asegura que, para tamaños muestrales grandes, la probabilidad de que la media muestral difiera significativamente de la media poblacional tiende a cero. Este resultado es fundamental para garantizar que la media muestral sea un estimador consistente de la media poblacional.

III .ii. Explica los conceptos de "distribución muestral" (sampling distribution) de un estimador y su "error estándar".

La distribución muestral de un estimador es la distribución de probabilidad que describe cómo varía ese estimador cuando se toman múltiples muestras aleatorias de un tamaño fijo n de una población.

- Es un concepto teórico que describe la variabilidad inherente de un estadístico debido al muestreo.
- La forma de la distribución muestral depende del estadístico, del tamaño muestral n, y de la distribución de la población original.

El error estándar de un estadístico es una medida de la variabilidad esperada del estadístico debido al muestreo aleatorio. Matemáticamente, es el desvío estándar de la distribución muestral del estadístico.

IV. Luego de leer el libro del Genesis un teólogo desea conocer la proporción de hombres y mujeres en una comunidad bíblica. Para ello, selecciona aleatoriamente 150 personas mencionadas en los relatos bíblicos y encuentra que 90 son hombres y 60 son mujeres.

El análisis se realiza utilizando una fórmula que calcula la proporción como el cociente entre el número de hombres observados y el tamaño de la muestra. Aplicando esta fórmula a los datos recolectados se obtiene un valor numérico de 0.6.

Pregunta:: Con base en la situación descrita, identifica

- **El estimando:** La proporción verdadera de hombres en la comunidad bíblica (desconocida).
- **El estimador:** La fórmula $\hat{p} = \frac{\text{Número de hombres observados}}{\text{Tamaño de la muestra}}$.
- **El estimado ($\hat{p} = 0.6$)**: El valor calculado de 0.6, basado en

V. El arqueólogo Dr. Ezekiel Ben-David, en sus estudios sobre registros históricos de Galilea en el año 34 d.C., recopiló datos sobre la altura y el sexo de individuos, posiblemente descendientes de los fundadores bíblicos Adán y Eva.

Sexo	Altura (cm)
Varón	150
Hembra	140
Hembra	147
Hembra	162
Varón	199
Varón	169
Varón	168
Hembra	168

Sexo	Altura (cm)
Hembra	152
Hembra	171
Hembra	159
Varón	167
Hembra	166
Varón	161
Hembra	166
Hembra	157
Varón	159
Hembra	154
Hembra	164
Varón	148

:

V .i. Estima la proporción de hombres y construye un intervalo de confianza al 95% para dicha proporción. Interpreta el resultado tanto desde una perspectiva estadística como sustantiva. Justifica tus decisiones metodológicas.

Estimación de la proporción: - Número de hombres: $n_{\text{hombres}} = 8$. - Tamaño total de la muestra: $n_{\text{total}} = 20$. - Proporción estimada de hombres:

$$\hat{p} = \frac{n_{\text{hombres}}}{n_{\text{total}}} = \frac{8}{20} = 0.4.$$

Intervalo de confianza al 95%: - Error estándar de la proporción:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_{\text{total}}}} = \sqrt{\frac{0.4(1-0.4)}{20}} \approx 0.1095.$$

- Usando un valor crítico $z = 1.96$ (para un nivel de confianza del 95%):

$$IC_{95\%} = \hat{p} \pm z \cdot SE(\hat{p}).$$

- Límite inferior:

$$0.4 - 1.96 \cdot 0.1095 \approx 0.1853.$$

- Límite superior:

$$0.4 + 1.96 \cdot 0.1095 \approx 0.6147.$$

- Intervalo de confianza:

$$IC_{95\%} = (0.1853, 0.6147).$$

Interpretación estadística: Estimamos que el 40% de los individuos son hombres. Con un 95% de confianza, la proporción verdadera de hombres en esta comunidad se encuentra entre 18.5% y 61.5%.

Interpretación sustantiva: La proporción estimada sugiere que en esta comunidad bíblica hay más mujeres que hombres. Sin embargo, dado el intervalo de confianza relativamente amplio, no podemos descartar proporciones más cercanas a la paridad debido al tamaño reducido de la muestra.

V.ii. Realiza un test de hipótesis para evaluar si la altura promedio de los hombres era mayor que la de las mujeres. Plantea la hipótesis nula y alternativa, determina el test estadístico apropiado y calcula su valor p. Interpreta el resultado tanto desde una perspectiva estadística como sustantiva. Justifica tus decisiones metodológicas.

Planteamiento de las hipótesis: - H_0: La altura promedio de los hombres es menor o igual que la de las mujeres ($\mu_{\text{hombres}} \leq \mu_{\text{mujeres}}$). - H_1: La altura promedio de los hombres es mayor que la de las mujeres ($\mu_{\text{hombres}} > \mu_{\text{mujeres}}$).

Datos resumidos: - Promedio de altura de los hombres ($\bar{x}_{\text{hombres}}$): 165.13 cm. - Promedio de altura de las mujeres ($\bar{x}_{\text{mujeres}}$): 158.83 cm. - Tamaño de muestra: $n_{\text{hombres}} = 8$, $n_{\text{mujeres}} = 12$.

Test estadístico: - Usamos una prueba t de comparación de medias con la alternativa unilateral ($\mu_{\text{hombres}} > \mu_{\text{mujeres}}$):

$$t = \frac{\bar{x}_{\text{hombres}} - \bar{x}_{\text{mujeres}}}{\sqrt{\frac{s^2_{\text{hombres}}}{n_{\text{hombres}}} + \frac{s^2_{\text{mujeres}}}{n_{\text{mujeres}}}}}.$$

- Valor observado de t: $t = 1.13$. - Valor p: $p \approx 0.137$.

Interpretación estadística: - Dado que $p = 0.137 > 0.05$, no rechazamos la hipótesis nula al nivel de significancia del 5%. Esto significa que no hay suficiente evidencia estadística para afirmar que la altura promedio de los hombres es mayor que la de las mujeres.

Interpretación sustantiva: - Aunque los hombres parecen ser más altos en promedio (165.13 cm) que las mujeres (158.83 cm), la diferencia no es lo suficientemente grande como para ser estadísticamente significativa, posiblemente debido al tamaño pequeño de las muestras y la variabilidad en los datos.

VI. En sus investigaciones, el arqueólogo Dr. Ezekiel Ben-David descubrió los resultados de un plebiscito popular llevado a cabo en comunidades rurales y urbanas de Galilea (año 34 d.C.), respecto a la adopción de un nuevo código moral con dos opciones: “Tradición” y “Cambio”.

La siguiente tabla de contingencia resume los resultados obtenidos:

	Rural	Urbano
“Tradición”	50	60
“Cambio”	50	140

VI.i. Realiza un test de χ^2 para evaluar si existe una asociación entre el tipo de comunidad (rural o urbana) y la preferencia por el nuevo código moral. Detalla los pasos involucrados en el cálculo del estadístico χ^2 y proporciona una interpretación sustantiva del resultado.

- **Hipótesis nula (H_0):** No existe asociación entre el tipo de comunidad (rural o urbana) y la preferencia por el código moral.
- **Hipótesis alternativa (H_a):** Existe asociación entre el tipo de comunidad y la preferencia por el código moral.
- **Datos observados:**

	Rural	Urbano
”Tradición”	50	60
”Cambio”	50	140

- **Frecuencias esperadas:**
  Utilizando la fórmula $E_{ij} = \frac{R_i \cdot C_j}{N}$, donde R_i y C_j son los totales de las filas y columnas, respectivamente:

	Rural	Urbano
”Tradición”	36.67	73.33
”Cambio”	63.33	126.67

- **Estadístico Chi-cuadrado y valor p:**

$$\chi^2 = 10.64, \quad p = 0.0011$$

- **Grados de libertad:** $df = (2-1)(2-1) = 1$

Interpretación:
El valor p (0.0011) es mucho menor que un nivel de significancia típico ($\alpha = 0.05$), por lo que rechazamos la hipótesis nula de independencia. Esto indica que existe evidencia de una asociación entre el tipo de comunidad (rural o urbana) y la preferencia por el nuevo código moral.

VI.ii. Calcula las proporciones muestrales que estiman las siguientes probabilidades:

- $P(\text{"Cambio"} \mid \text{Rural}) = \hat{p}_{\text{R}} = \frac{50}{100} = 0.5$
- $P(\text{"Cambio"} \mid \text{Urbano}) = \hat{p}_{\text{U}} = \frac{140}{200} = 0.7$

VI.iii. Realiza un test de diferencia en proporciones para evaluar la hipótesis:

- **Hipótesis nula (H_0):** $P(\text{"Cambio"} \mid \text{Rural}) = P(\text{"Cambio"} \mid \text{Urbano})$

- **Hipótesis alternativa (H_a):** $P(\text{"Cambio"} \mid \text{Rural}) < P(\text{"Cambio"} \mid \text{Urbano})$
- **Estadístico de prueba:**
  Usando la fórmula:
  $$z = \frac{\hat{p}_{\text{R}} - \hat{p}_{\text{U}}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_{\text{R}}} + \frac{1}{n_{\text{U}}}\right)}}$$
  donde $\hat{p}$ es la proporción combinada:
  $$\hat{p} = \frac{50+140}{100+200} = 0.6333$$
  Sustituyendo valores:
  $$z = \frac{0.5 - 0.7}{\sqrt{0.6333 \cdot (1 - 0.6333)\left(\frac{1}{100} + \frac{1}{200}\right)}} = -3.39$$
- **Valor p:**
  Para una prueba unilateral:
  $$p = \mathrm{P}(Z < -3.39) \approx 0.00035$$

Interpretación:
El valor p es muy pequeño ($p < 0.05$), por lo que rechazamos H_0. Esto proporciona evidencia sólida de que la proporción de preferencia por "Cambio" es significativamente mayor en comunidades urbanas que en comunidades rurales.

VI. Galileo Galilei y la Ley de la Caída Libre

Galileo Galilei descubrió que la distancia recorrida por un objeto en caída libre desde el reposo está dada por la ecuación:

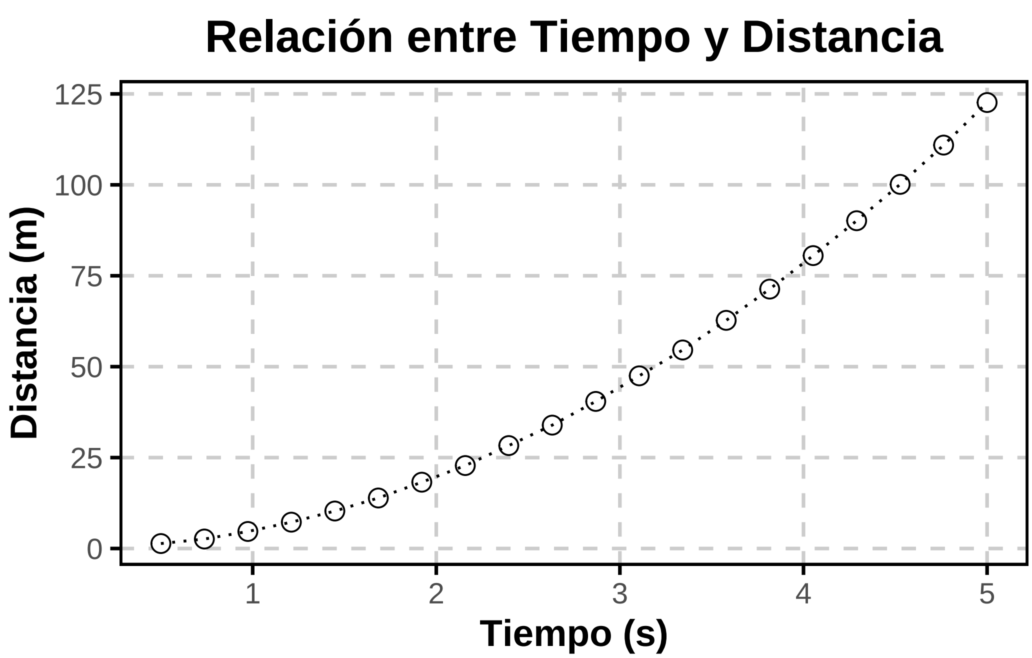
$$d = 4.9t^2$$

Donde: - (d) es la distancia recorrida (en metros). - (t) es el tiempo transcurrido (en segundos).

Un físico decidió poner a prueba esta ley realizando un experimento con un tamaño muestral de ($n = 20$). Midió el tiempo y la distancia recorrida por un objeto en caída libre en condiciones reales, por lo que los datos presentan un pequeño ruido aleatorio. Los resultados se resumen en la siguiente tabla:

Variable	Media	Desviación estándar (SD)	Covarianza (t, d)
Tiempo (t)	2.75 s	1.40 s	52.83
Distancia (d)	46.18 m	38.62 m	-

El siguiente gráfico muestra la relación entre el tiempo y la distancia reportada en el experimento:



VI.i. ¿Qué tipo de relación observas entre el tiempo (t) y la distancia (d) según los datos?

La relación entre el tiempo y la distancia en caída libre muestra una **dependencia cuadrática**. Según la fórmula teórica $d = 4.9t^2$, la distancia aumenta con el cuadrado del tiempo. En el gráfico proporcionado, se observa una relación creciente y no lineal, lo que concuerda con la ecuación descrita.

VI.ii. Calcula el coeficiente de correlación de Pearson entre el tiempo (t) y la distancia (d). Comenta sobre el signo y la magnitud de la correlación obtenida.

El coeficiente de correlación de Pearson (r) mide la fuerza y dirección de la relación lineal entre t y d. Se calcula como:

$$r = \frac{\mathrm{Cov}(t, d)}{\sigma_t \cdot \sigma_d},$$

donde: - $\mathrm{Cov}(t, d) = 52.83$ (covarianza), - $\sigma_t = 1.40$ (desviación estándar de t), - $\sigma_d = 38.62$ (desviación estándar de d).

Sustituyendo los valores:

$$r = \frac{52.83}{1.40 \cdot 38.62} \approx 0.977.$$

$r \approx 0.977$ indica una correlación muy fuerte. El signo positivo implica que a medida que t aumenta, d también aumenta.

VI.iii. ¿Crees que es una medida adecuada para describir esta relación?

El coeficiente de correlación de Pearson mide la relación lineal entre t y d, pero la relación real entre estas variables es **no lineal (cuadrática)**. Aunque r es alto debido a la monotonía de la relación (ambas variables aumentan juntas), no captura adecuadamente la naturaleza no lineal de los datos. Por tanto, a pesar de que r proporciona información sobre la fuerza de la relación creciente, no describe completamente la forma cuadrática subyacente.