

# Análisis de Investigación

---

## Proyecto de Aplicación Parcial I

### ESTADISTICA

Isaac Oña

Variable de estudio:

En los últimos seis meses, ¿cuánto dinero aproximadamente ha gastado en recursos para usar ChatGPT?



**En los últimos seis meses, ¿cuánto dinero aproximadamente ha gastado en recursos para usar ChatGPT?**

En el ámbito académico, la adopción de herramientas tecnológicas puede variar considerablemente entre diferentes disciplinas. La Universidad de las Fuerzas Armadas - ESPE busca comprender mejor cómo sus estudiantes de las carreras de Software, Mercadotecnia y Biotecnología utilizan herramientas innovadoras como ChatGPT en sus actividades académicas y personales. Esta encuesta se enfoca en identificar el nivel de uso y los recursos económicos invertidos en esta tecnología.

### **Datos no Agrupados**

A continuación, se mostrará las fórmulas ocupadas en Datos no Agrupados

### **Medidas de Tendencia Central**

#### **Media**

$$Media = \frac{\sum x}{n}$$

#### **Mediana**

$$Me = \frac{n + 1}{2}$$

#### **Moda**

$$Mo = \#Que \text{ más se repite}$$

### **Medidas de Dispersión**

#### **Amplitud**

$$Amplitud = Valor \text{ mayor} - Valor \text{ Menor.}$$

#### **Desviación Media Absoluta**

$$DAM = \frac{\sum (xi - x)}{n}$$

### Varianza

$$s^2 = \frac{\sum (xi - x)^2}{n - 1}$$

### Desviación Estándar

$$s = \sqrt{\frac{\sum f(xi - x)^2}{n - 1}}$$

### Otras medidas de posición

#### Cuartiles

$$Q_1 = \frac{(n + 1)(1)}{4}$$

$$Q_2 = \frac{(n + 1)(2)}{4}$$

$$Q_3 = \frac{(n + 1)(3)}{4}$$

#### Curtosis

$$CA = \frac{3(X - Me)}{s}$$

#### Coefficiente de variación

$$CV = \frac{s}{X} \cdot 100$$

## TABLA DE DATOS NO AGRUPADOS

Nº	Datos Ordenados	$\bar{x}$	$ x - \bar{x} $	$ x - \bar{x} ^2$	$x^2$
1	0.00	7.77	7.77	60.44	0.00
2	0.00	7.77	7.77	60.44	0.00
3	0.00	7.77	7.77	60.44	0.00
4	0.00	7.77	7.77	60.44	0.00
5	0.00	7.77	7.77	60.44	0.00
6	0.00	7.77	7.77	60.44	0.00
7	0.00	7.77	7.77	60.44	0.00
8	0.00	7.77	7.77	60.44	0.00
9	0.00	7.77	7.77	60.44	0.00
10	0.00	7.77	7.77	60.44	0.00
11	0.00	7.77	7.77	60.44	0.00
12	0.00	7.77	7.77	60.44	0.00
13	0.00	7.77	7.77	60.44	0.00
14	0.00	7.77	7.77	60.44	0.00
15	0.00	7.77	7.77	60.44	0.00
16	0.00	7.77	7.77	60.44	0.00
17	0.00	7.77	7.77	60.44	0.00
18	0.00	7.77	7.77	60.44	0.00
19	0.00	7.77	7.77	60.44	0.00
20	0.00	7.77	7.77	60.44	0.00
21	0.00	7.77	7.77	60.44	0.00
22	0.00	7.77	7.77	60.44	0.00
23	0.00	7.77	7.77	60.44	0.00
24	0.00	7.77	7.77	60.44	0.00
25	0.00	7.77	7.77	60.44	0.00
26	0.00	7.77	7.77	60.44	0.00
27	0.00	7.77	7.77	60.44	0.00
28	0.00	7.77	7.77	60.44	0.00
29	0.00	7.77	7.77	60.44	0.00
30	0.00	7.77	7.77	60.44	0.00
31	0.00	7.77	7.77	60.44	0.00
32	0.00	7.77	7.77	60.44	0.00
33	0.00	7.77	7.77	60.44	0.00
34	0.00	7.77	7.77	60.44	0.00
35	0.00	7.77	7.77	60.44	0.00
36	0.00	7.77	7.77	60.44	0.00
37	0.00	7.77	7.77	60.44	0.00
38	0.00	7.77	7.77	60.44	0.00
39	0.00	7.77	7.77	60.44	0.00
40	0.00	7.77	7.77	60.44	0.00
41	0.00	7.77	7.77	60.44	0.00
42	0.00	7.77	7.77	60.44	0.00
43	0.00	7.77	7.77	60.44	0.00
44	0.00	7.77	7.77	60.44	0.00
45	0.00	7.77	7.77	60.44	0.00
46	0.00	7.77	7.77	60.44	0.00
47	0.00	7.77	7.77	60.44	0.00
48	0.00	7.77	7.77	60.44	0.00
49	0.00	7.77	7.77	60.44	0.00
50	0.00	7.77	7.77	60.44	0.00
51	0.00	7.77	7.77	60.44	0.00
52	0.00	7.77	7.77	60.44	0.00
53	0.00	7.77	7.77	60.44	0.00
54	0.00	7.77	7.77	60.44	0.00
55	0.00	7.77	7.77	60.44	0.00
56	0.00	7.77	7.77	60.44	0.00
57	0.00	7.77	7.77	60.44	0.00
58	0.00	7.77	7.77	60.44	0.00
59	0.00	7.77	7.77	60.44	0.00
60	0.00	7.77	7.77	60.44	0.00
61	0.00	7.77	7.77	60.44	0.00
62	0.00	7.77	7.77	60.44	0.00
63	0.00	7.77	7.77	60.44	0.00
64	0.00	7.77	7.77	60.44	0.00
65	0.00	7.77	7.77	60.44	0.00
66	0.00	7.77	7.77	60.44	0.00
67	0.00	7.77	7.77	60.44	0.00
68	0.00	7.77	7.77	60.44	0.00
69	0.00	7.77	7.77	60.44	0.00
70	0.00	7.77	7.77	60.44	0.00

71	0.00	7.77	7.77	60.44	0.00	106	20.00	7.77	12.23	149.46	400.00
72	0.00	7.77	7.77	60.44	0.00	107	20.00	7.77	12.23	149.46	400.00
73	0.00	7.77	7.77	60.44	0.00	108	20.00	7.77	12.23	149.46	400.00
74	0.00	7.77	7.77	60.44	0.00	109	20.00	7.77	12.23	149.46	400.00
75	1.20	7.77	6.57	43.22	1.44	110	20.00	7.77	12.23	149.46	400.00
76	2.00	7.77	5.77	33.34	4.00	111	20.00	7.77	12.23	149.46	400.00
77	2.00	7.77	5.77	33.34	4.00	112	20.00	7.77	12.23	149.46	400.00
78	2.00	7.77	5.77	33.34	4.00	113	20.00	7.77	12.23	149.46	400.00
79	2.00	7.77	5.77	33.34	4.00	114	20.00	7.77	12.23	149.46	400.00
80	2.30	7.77	5.47	29.97	5.29	115	20.00	7.77	12.23	149.46	400.00
81	3.00	7.77	4.77	22.80	9.00	116	20.50	7.77	12.73	161.94	420.25
82	3.00	7.77	4.77	22.80	9.00	117	22.30	7.77	14.53	210.99	497.29
83	3.00	7.77	4.77	22.80	9.00	118	22.30	7.77	14.53	210.99	497.29
84	3.00	7.77	4.77	22.80	9.00	119	22.40	7.77	14.63	213.91	501.76
85	3.00	7.77	4.77	22.80	9.00	120	23.00	7.77	15.23	231.82	529.00
86	3.20	7.77	4.57	20.93	10.24	121	25.00	7.77	17.23	296.72	625.00
87	3.50	7.77	4.27	18.27	12.25	122	25.00	7.77	17.23	296.72	625.00
88	3.85	7.77	3.92	15.40	14.82	123	25.00	7.77	17.23	296.72	625.00
89	4.00	7.77	3.77	14.25	16.00	124	26.30	7.77	18.53	343.20	691.69
90	4.20	7.77	3.57	12.78	17.64	125	30.00	7.77	22.23	493.98	900.00
91	5.00	7.77	2.77	7.70	25.00	126	30.00	7.77	22.23	493.98	900.00
92	5.00	7.77	2.77	7.70	25.00	127	30.20	7.77	22.43	502.91	912.04
93	5.00	7.77	2.77	7.70	25.00	128	30.20	7.77	22.43	502.91	912.04
94	5.00	7.77	2.77	7.70	25.00	129	35.40	7.77	27.63	763.17	1253.16
95	10.00	7.77	2.23	4.95	100.00	130	40.00	7.77	32.23	1038.49	1600.00
96	10.00	7.77	2.23	4.95	100.00	131	40.00	7.77	32.23	1038.49	1600.00
97	10.00	7.77	2.23	4.95	100.00	132	40.60	7.77	32.83	1077.52	1648.36
98	10.00	7.77	2.23	4.95	100.00	133	55.40	7.77	47.63	2268.19	3069.16
99	10.20	7.77	2.43	5.88	104.04	134	56.20	7.77	48.43	2345.03	3158.44
100	10.20	7.77	2.43	5.88	104.04	135	63.20	7.77	55.43	3071.99	3994.24
101	10.20	7.77	2.43	5.88	104.04	Σ	7.77		1331.10	22415.14	30574.81
102	10.20	7.77	2.43	5.88	104.04						
103	10.50	7.77	2.73	7.43	110.25						
104	15.00	7.77	7.23	52.21	225.00						
105	15.00	7.77	7.23	52.21	225.00						
106	20.00	7.77	12.23	149.46	400.00						

**Tabla 1 DATOS NO AGRUPADOS**

Estos datos fueron analizados utilizando Microsoft Excel, una herramienta que facilita el cálculo de medidas estadísticas.

Media	7.77
Moda	0.00
Mediana	0.00
Amplitud	63.20
DAM	9.86
Varianza	167.28
Desviación Estándar	12.93

**Figura 2 Medidas de tendencia central**

La tabla presenta diversas medidas estadísticas descriptivas de un conjunto de datos, proporcionando una visión clara de su distribución y variabilidad. La media (7.77) representa el

valor promedio de los datos, mientras que la moda y la mediana son ambas 0, indicando que la mayoría de los datos son 0 o menores. La amplitud (63.20) mide la diferencia entre el valor más alto y el más bajo. La desviación absoluta media (9.86), la varianza (167.28) y la desviación estándar (12.93) cuantifican la dispersión de los datos respecto a la media, con la desviación estándar siendo la raíz cuadrada de la varianza. En conjunto, estas medidas ofrecen una comprensión integral de cómo se distribuyen y varían los datos en este conjunto específico.

### **Datos Agrupados.**

#### **Medidas de Tendencia Central**

##### **Media**

$$Media = \frac{Fx}{n}$$

##### **Mediana**

$$Me = Li + \frac{(\frac{n}{2} - FA \uparrow)}{f} * Ai$$

##### **Moda**

$$Mo = Li + \left[ \frac{d1}{(d1 + d2)} * Ai \right]$$

#### **Medidas de Dispersión**

##### **Amplitud**

$$Amplitud = Valor\ mayor - Valor\ menor$$

##### **Desviación Media Absoluta**

$$DAM = \frac{\sum f(xi - x)}{n}$$



avanzados y visualizar los resultados de manera clara. Con esta aplicación, se verificaron y complementaron los resultados obtenidos en Excel.

Descriptivas	
	Datos Ordenados
N	135
Perdidos	0
Media	7.77
Mediana	0.00
Moda	0.00
Suma	1050
Desviación estándar	12.9
Varianza	167
Mínimo	0.00
Máximo	63.2
Curtosis	4.21
Error est. curtosis	0.414

### **Probabilidad.**

#### **Reglas de probabilidad.**

De un total de 135 estudiantes que usan ChatGPT, tenemos la siguiente distribución:

- 30 estudiantes son de Biotecnología.
- 50 estudiantes son de Software.
- 20 estudiantes son de Mercadotecnia.
- 15 estudiantes son de Software y Mercadotecnia.
- 25 estudiantes son de Biotecnología y Software.

**Determine la probabilidad de que, al escoger una persona al azar, esta gaste más de 10\$, en usar ChatGPT al mes.**



**Eventos:**

**Evento B:** El estudiante gasta más de 10 USD al mes y pertenece a Biotecnología.

**Evento S:** El estudiante gasta más de 10 USD al mes y pertenece a Software.

**Evento SyM:** El estudiante gasta más de 10 USD al mes y pertenece a Software y Mercadotecnia.

$$P(B) = \frac{30}{135}$$

$$P(S) = \frac{50}{135}$$

$$P(M) = \frac{20}{135}$$

$$P(SoM) = \frac{15}{135}$$

$$P(BoS) = \frac{25}{135}$$

Calcule la probabilidad de que un estudiante gaste más \$10, sea de Software o Biotecnología:

$$P(BoS) = P(B) + P(S) - P(BoS)$$

$$P(BoS) = \frac{30}{135} + \frac{50}{135} - \frac{25}{135}$$

$$P(BoS) = \frac{11}{27} \approx 0.407 \approx 40.74\%$$

Respuesta: La probabilidad de que, al escoger una persona al azar, ésta sea de Software o de Biotecnología y use ChatGPT y paguen \$10 es aproximadamente **40.74 %**.

### **Teorema de Bayes.**

En una universidad, se sabe que los estudiantes que usan ChatGPT se distribuyen de la siguiente manera:

- El **29.63%** de los estudiantes pertenecen a Biotecnología.
- El **33.33%** de los estudiantes pertenecen a Software.
- El **37.04%** de los estudiantes pertenecen a Mercadotecnia.

Esto se traduce en:

- **Estudiantes de Biotecnología:** 40.
- **Estudiantes de Software:** 45.
- **Estudiantes de Mercadotecnia:** 50.

Además:

- El **80%** de los estudiantes de Software han gastado más de \$10 en ChatGPT.
- El **50%** de los estudiantes de Biotecnología han gastado más de \$10 en ChatGPT.
- El **30%** de los estudiantes de Mercadotecnia han gastado más de \$10 en ChatGPT.

Si se selecciona un estudiante al azar que ha gastado más de \$10 en ChatGPT, ¿cuál es la probabilidad de que pertenezca al departamento de Biotecnología?

El Teorema de Bayes establece:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

A: El estudiante pertenece a Biotecnología.

B: El estudiante ha gastado más de \$10 en ChatGPT.

### 1. Probabilidades iniciales

Probabilidad de  $P(A)$ ,  $P(B)$  y  $P(C)$ :

$$P(A) = \frac{40}{135} \approx 0.2963$$

$$P(B) = \frac{45}{135} \approx 0.3333$$

$$P(C) = \frac{50}{135} \approx 0.3704$$

Probabilidades condicionales de gastar más de \$10  $P(B|A)$ ,  $P(B|B)$ ,  $P(B|C)$

$P(B|A) = 0.50$  (50% de los estudiantes de Biotecnología gastan más de \$10).

$P(B|B) = 0.80$  (80% de los estudiantes de Software gastan más de \$10).

$P(B|C) = 0.30$  (30% de los estudiantes de Mercadotecnia gastan más de \$10).

### 2. Total, de estudiantes que gastan más de \$10 ( $P(B)$ )

$$P(B) = P(B|A) \cdot P(A) + P(B|B) \cdot P(B) + P(B|C)$$

$$P(B) = (0.50 \cdot 0.2963) + (0.80 \cdot 0.3333) + (0.30 \cdot 0.3704)$$

$$P(B) = 0.1481 + 0.2667 + 0.1111 = 0.5259$$

$$P(B) \approx 0.5259$$

### 3. Aplicamos Bayes.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(A|B) = \frac{(0.50) \cdot (0.2963)}{0.5259}$$

$$P(A|B) = \frac{0.1481}{0.5259} \approx 0.2815$$

R= La probabilidad de que un estudiante que ha gastado más de \$10 en ChatGPT pertenezca al departamento de Biotecnología es: **28.15%**

### Distribución Binomial

Supongamos que el 35% de los estudiantes de Software gastan mas de 10 \$ en ChatGPT. Se selecciona una muestra de 8 estudiantes de Software al azar. Queremos calcular la probabilidad de que exactamente 5 de estos estudiantes hayan gastado mas de \$10 en ChatGPT.

Datos:

**(n): 8**

**(p): 0.35**

**(k): 5**

La distribución binomial se describe por la fórmula:

$$p(k) = C_K^n p^k q^{(n-k)}$$

$$[P(X = 5) = \binom{8}{5} (0.35)^5 (0.65)^3]$$

$$(0.35)^5 = 0.0052521875, \quad (0.65)^3 = 0.274625$$

$$P(X = 5) = 56 \cdot 0.0052521875 \cdot 0.274625$$

$$P(X = 5) \approx 0.08095$$

$$P(X = 5) = 0.081 \approx 8.1\%$$

R= La probabilidad de que exactamente 5 de los 8 estudiantes seleccionados hayan gastado más de \$10 en ChatGPT es aproximadamente 08.1%

### **Distribución Poisson.**

Supongamos que el número promedio de estudiantes que gastan más de \$10 en ChatGPT en un día es  $\mu = 4$ . Queremos calcular la probabilidad de que exactamente 6 estudiantes gasten más de \$10 en un día.

La distribución de Poisson se define con la fórmula:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

Donde:

$\mu=4$

$k=6$

-Sustituimos los valores:

$$P(6) = \frac{4096 \cdot 0.01831564}{720}$$

$$P(6) \approx 0.1048$$

$$P(6) = 10.5\%.$$

R= La probabilidad de que exactamente 6 estudiantes gasten más de \$10 en un día es aproximadamente **0.105** o **10.5%**.

### **Distribución Hipergeométrica**

Supongamos que en una clase hay un total de 25 estudiantes de Biotecnología, 20 de Software y 10 de Mercadotecnia, y todos usan ChatGPT. Si seleccionamos una muestra de 12 estudiantes al azar, ¿cuál es la probabilidad de que exactamente 5 hayan gastado más de \$10 usando ChatGPT y sean de Biotecnología?

La fórmula para la distribución hipergeométrica es:

$$P(x) = \frac{n - sC_{n-x} * sC_x}{Nc_n}$$

Donde:

$$N=25+20+10=55 \quad N = 25 + 20 + 10 = 55$$

$$K=25 \quad K = 25$$

$$n=12 \quad n = 12$$

$$k=5$$

Por lo tanto:

$$P(5) = \frac{C_5^{25} * C_7^{30}}{C_{12}^{55}}$$

$$P(5) = 0.3676 \approx 36.76\%$$

R= La probabilidad de que exactamente 5 estudiantes de Biotecnología hayan gastado más de \$10 usando ChatGPT en una muestra de 12 estudiantes es aproximadamente 0.3676, o 36.76%.

### **Distribución Uniforme**

Supongamos que en una clase, el gasto en ChatGPT por los estudiantes varía uniformemente entre \$5 y \$50. Queremos calcular la probabilidad de que un estudiante haya gastado entre \$20 y \$35.

#### **Datos del problema:**

a=5 (gasto mínimo en dólares)

b=50 (gasto máximo en dólares)

Queremos calcular  $P(20 \leq X \leq 35)$

#### **Por lo tanto:**

$$P(20 \leq X \leq 35) = \frac{35 - 20}{50 - 5} = \frac{15}{45} = \frac{1}{3} \approx 33.33\%$$

R= La probabilidad de que un estudiante haya gastado entre \$20 y \$35 es aproximadamente 33.33%.

### Distribución Normal.

Supongamos que el gasto promedio en ChatGPT por los estudiantes de una clase sigue una distribución normal con una media ( $\mu$ ) de \$30 y una desviación estándar ( $\sigma$ ) de \$10. Queremos calcular la probabilidad de que un estudiante haya gastado entre \$20 y \$40.

#### Datos del problema:

$\mu = 30$  (media del gasto en dólares)

$\sigma = 10$  (desviación estándar en dólares)

Queremos calcular  $P(20 < x < 40)$

Para nuestro Caso:

Para  $x=20$

$$Z = \frac{20 - 30}{10} = \frac{-10}{10} = -1$$

Para  $x=40$

$$Z = \frac{40 - 30}{10} = \frac{10}{10} = 1$$

Usamos la tabla Z para encontrar las probabilidades correspondientes a  $Z=-1$  y  $Z = 1$

$$P(Z \leq -1) \approx 0.1587$$

$$P(Z \leq 1) \approx 0.8413$$



La probabilidad de que  $X$  esté entre 20 y 40 es:

$$P(20 \leq X \leq 40) = P(Z \leq 1) - P(Z \leq -1)$$

$$P(20 \leq X \leq 40) = 0.8413 - 0.1587 = 0.6826$$

R= La probabilidad de que un estudiante haya gastado entre \$20 y \$40 es aproximadamente 0.6826, o 68.26%.

### **Distribución Exponencial.**

Supongamos que el tiempo que un estudiante tarda en gastar más de \$10 en ChatGPT sigue una distribución exponencial con una tasa promedio de 1 vez por cada 5 días. Queremos calcular la probabilidad de que un estudiante gaste más de \$10 en un periodo entre 3 y 7 días.

#### **Datos del problema:**

$\lambda=15$  (tasa de ocurrencia, eventos por día)

Queremos calcular  $P(3 \leq X \leq 7)$

Por lo tanto:

$$F(x) = 1 - e^{-\lambda x} \text{ para } x \geq 0$$

$$F(x) = 1 - e^{-0.2x}$$

Calculamos  $F(b)$  y  $F(a)$ :

$$F(7) = 1 - e^{-0.2 \cdot 7} = 1 - e^{-1.4} \approx 1 - 0.2466 = 0.7534$$

$$F(3) = 1 - e^{-0.2 \cdot 3} = 1 - e^{-0.6} \approx 1 - 0.5488 = 0.4512$$

La probabilidad de que X esté entre 3 y 7 días es:

$$P(3 \leq X \leq 7) = F(7) - F(3)$$

$$P(3 \leq X \leq 7) = 0.7534 - 0.4512 = 0.3022$$

R = La probabilidad de que un estudiante gaste más de \$10 en un periodo entre 3 y 7 días es aproximadamente 0.3022 o 30.22%.

### Referencias:

The jamovi project (2024). *jamovi*. (Version 2.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.

R Core Team (2024). *R: A Language and environment for statistical computing*. (Version 4.4) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2024-08-07).



