RAZONAMIENTO ESTADÍSTICO INFERENCIAL

La estadística Inferencial, es el proceso por el cual se deducen (infieren) propiedades o características de una población a partir de una muestra significativa.

	Población	**Muestra**
Definición	Colección de elementos considerados	Parte o porción de la población seleccionada para su estudio
Características	“Parámetros”	“Estadísticos”
Símbolos	Tamaño de la población = N	Tamaño de la muestra = n
	Media de la población = μ	Media de la muestra = $\bar{x}$
	Desviación estándar de la población = σ	Desviación estándar de la muestra = s

Método de muestreo

- **Métodos no probabilísticos:** Interviene la opinión del investigador para obtener cada elemento de la muestra.

- **Métodos probabilísticos:** Muestra que se selecciona de modo que cada integrante de la población en estudio tenga una probabilidad conocida (pero distinta de cero) de ser incluido en la muestra.
  - Muestreo Aleatorio Simple
  - Muestreo Aleatorio Sistemático
  - Muestreo Aleatorio Estratificado
  - Muestreo Aleatorio Por Conglomerado

Muestreo aleatorio simple
Muestra seleccionada de manera que cada integrante de la población tenga la misma probabilidad de quedar incluido.

Ejemplo: un bingo, introduzco los números en un ánfora y selecciono una muestra al azar

Muestreo aleatorio sistemático
Los integrantes o elementos de la población se ordenan en alguna forma (Ejemplo: alfabéticamente) se selecciona al azar un punto de partida y después se elige para la muestra cada k-ésimo elemento de la población.

Ejemplo: se desea establecer una muestra 100 empleados de los 3000 que tiene una empresa, para lo cual ordeno alfabéticamente a los empleados, divido 3000/100 = 30 y selecciona a uno de cada treinta empleados

Muestreo aleatorio estratificado
Una población se divide en subgrupos denominados estratos y se selecciona una muestra de cada uno

Estrato	**Edades**	**N.º de empleados**	**% del total**	**Cantidad muestreada**
1 2 3 4 5	MENOS DE 25 AÑOS 26-30 AÑOS 31-35 AÑOS 36-40 AÑOS MÁS DE 41 AÑOS	8 35 189 115 5	2 10 54 33 1	1 5 27 16 1
TOTAL		**352**	**100**	**50**

Muestreo aleatorio por conglomerado
Se divide a la población en estratos (subunidades) se selecciona con que subunidades se va a trabajar y de las unidades seleccionadas, se toma una muestra aleatoriamente

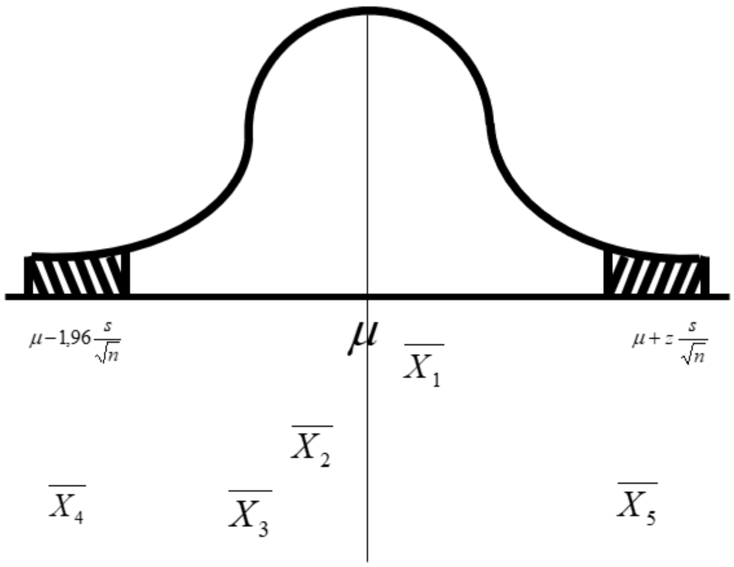
Ejemplo IPC: Guayaquil, Machala, Portoviejo, Quito, Ambato, Cuenca y Manta, Esmeraldas y Quevedo, Riobamba, Loja y Latacunga. Con estas ciudades se cubre el 67% de la población urbana del país,

Conceptos iniciales
Estimación Puntual: Estadístico calculado a partir de la información obtenida de la muestra y que se usa para estimar el parámetro poblacional

Intervalo de confianza: es un conjunto de valores obtenido a partir de los datos muestrales en el que hay una determinada probabilidad de que se encuentre el parámetro, a esta probabilidad se le conoce como el nivel de significancia

Error de muestreo: Diferencia entre un valor estadístico de muestra y su parámetro de población correspondiente

Intervalos de confianza



Intervalo de confianza para muestras mayores a 30 elementos

$$\overline{X} \pm z\frac{s}{\sqrt{n}}$$

Intervalo de confianza para muestras menores a 30 elementos

$$\overline{X} \pm t\frac{s}{\sqrt{n}}$$

Desviación estándar

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}} \qquad s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}}$$

Proporciones

Proporción: Fracción, razón o porcentaje que indica la parte de la muestra o población que tiene una característica determinada

Proporción muestral: $p = \frac{x}{n}$

Intervalos de confianza para una proporción poblacional

$$p \pm z\sigma_p$$

Error estándar de la proporción muestral

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Ejercicio:
Suponga que se toma una muestra de 30 empleados de los cuales reciben en promedio $349 y una desviación estándar de $110. ¿Cuál es el intervalo de confianza?

$\overline{X} \pm z\frac{s}{\sqrt{n}}$ $\qquad 349 \pm 39$

$349 \pm 1{,}96\frac{110}{\sqrt{30}}$ $\qquad 310 \pm 389$

Suponga que se toma una muestra de 20 empleados de los cuales reciben en promedio $346 y una desviación estándar de $126. ¿Cuál es el intervalo de confianza?

$\overline{X} \pm t\frac{s}{\sqrt{n}}$ $\qquad 346 \pm 59$

$346 \pm 2{,}093\frac{126}{\sqrt{20}}$ $\qquad 287 \pm 405$

Ejemplo: Proporciones
En una muestra aleatoria de 2000 miembros de sindicato, se tiene que 1600 están a favor de fusionarse con otra empresa ¿Cuál es el valor estimado de la proporción poblacional? ¿Cuál es el intervalo de confianza al 95% de confianza?

$$p = \frac{x}{n}$$

$$p = \frac{1600}{2000} = 0{,}80$$

$$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

$$0{,}80 \pm 1{,}96\sqrt{\frac{0{,}80(1-0{,}80)}{2000}} = 0{,}80 \pm 0{,}018$$

Pruebas de hipótesis para una muestra

Hipótesis: Es una afirmación sobre una población, que puede someterse a pruebas al extraer una muestra aleatoria.

Prueba de hipótesis: Formular una teoría y luego contrastarla

Pasos para probar una hipótesis:

1. Prueba de hipótesis
2. Seleccionar el nivel de significancia
3. Calcular el valor estadístico de prueba
4. Formular la regla de decisión
5. Decidir

Paso 1. Plantear $\mathbf{H_0}$ Y $\mathbf{H_1}$

$$H_0: \mu = \mu_0$$
$$H_0: \mu > \mu_0$$

Hipótesis nula: Afirmación acerca del valor de un parámetro poblacional

$$H_1: \mu \neq \mu_0$$
$$H_1: \mu < \mu_0$$

Hipótesis Alternativa: Afirmación que se aceptará si los datos muestrales aseguran que es falsa

Paso 2. Seleccionar el nivel de significancia

Generalmente son del 5% o 1% (Error de tipo I y Error de tipo II)

- Error de tipo I: Rechazar la hipótesis nula, H0 cuando es verdadera

- Error de tipo II: Aceptar la hipótesis nula, H0 cuando es Falsa

Paso 3. Calcular el valor estadístico de prueba.

Estadísticos de pruebas como: Z, t de Student, F y Ji cuadrado

$z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ Para muestras grandes

$Z = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ Para proporciones

$t = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ Para muestras pequeñas

Paso 4. Formular la regla de decisión

Son las condiciones según las que se acepta o rechaza la hipótesis nula

Paso 5: Tomar una decisión
El valor observado de la estadística muestral se compara con el valor de estadística de prueba

Ejemplo: Prueba de hipótesis
La producción diaria en una planta industrial registrada durante n=30 días tiene una media Muestral de 990 toneladas y una desviación estándar de 20 toneladas, pruebe la hipótesis de que el promedio de la producción diaria difiere de 1000 toneladas por día.

Paso 1: Establecer hipótesis
$H0: \mu = 1000 toneladas$
$H1: \mu \neq 1000 toneladas$

Paso 2: Nivel de significancia (0.05%)
Paso 3: Valor estadístico de prueba
$\overline{x} = 990 toneladas$

$$n = 30días$$
$$\mu_0 = 1000toneladas$$
$$\sigma = 20toneladas$$
$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
$$z = \frac{990 - 1000}{\frac{20}{\sqrt{30}}} = -2,7$$

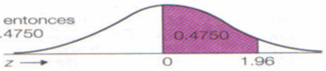
Áreas bajo la curva normal

DOS COLAS (0.05%)

0.05/2=0.025

0.50-0.025 =0.475

Ejemplo:
Si z = 1.96, entonces
P(0 a z) = 0.4750



0.50

UNA COLA

0.5-0.05=0.45

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	$\boxed{0.4750}$	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

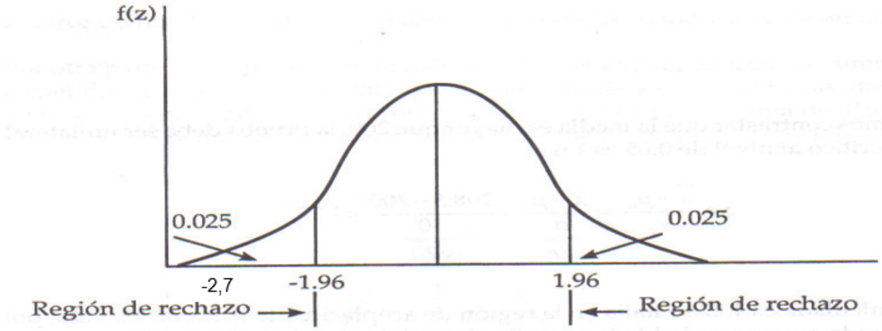
Valores críticos de Z en pruebas de hipótesis.

Valores de ∞	0,1	0,05	0,01	0,005	0,002
Z crítico para pruebas unilaterales	±1,28	±1,64	±2,33	±2,58	±2,88
Z crítico para pruebas bilaterales	±1,64	±1,96	±2,58	±2,81	±3,08

Paso 4: Formular la regla de decisión

Para un nivel de significancia de 0.05, la región de rechazo es z >1.96 o z< -1.96.

Paso 5: Tomar una decisión



Localización de la región de rechazo del ejemplo

Se rechaza H_0 μ no es igual a 1000 toneladas

Ejemplo: Prueba de hipótesis

El gerente de ventas de una empresa editora de libros afirma que cada uno de sus representantes realiza 40 visitas por semana a profesores. Varios vendedores dicen que esa estimación es muy baja. Para investigar lo anterior, una muestra aleatoria de 28 representantes de ventas reveló que el número medio de visitas realizadas la semana pasada fue de 42. Se calculó que la desviación estándar de la muestra fue de 2.1 visitas. Al nivel de significancia de 0.05, ¿se puede concluir que el número medio de visitas realizadas por vendedor y por semana es mayor que 40?

Paso 1: Establecer hipótesis

$H_0: \mu \angle 40$

$H_1: \mu > 40$

Paso 2: Nivel de significancia (0.05)

Paso 3: Estadístico de prueba

En este caso es T de student

$\mu = 40 visitas$

$\overline{x} = 42 visitas$

$s = 2.1 visitas$

$n = 28$

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{42 - 40}{\frac{2.1}{28}} = 5.04$$

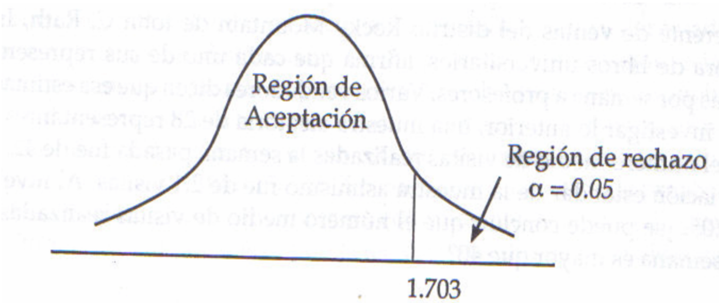
Ejemplo: con $gl = 9$ y área = 0.10 en la cola superior, $t = 1.383$

	Intervalos de confianza					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia para pruebas de una cola					
gl	0.100	0.050	0.025	0.010	0.005	0.0005
	Nivel de significancia para pruebas de dos colas					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
$\boxed{27}$	1.314	$\boxed{1.703}$	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

GRADOS DE LIBERTAD =28-1=27 VC = 1.703

Paso 4: Regla de decisión
Rechazo H0 SI t calculado es mayor a 1.703

Paso 5: Tomar decisión



T calculado = 5.04 cae en la región de rechazo. Por lo tanto, rechazamos H0. El número medio de visitas realizadas por vendedor y por semana es mayor que 40Prueba chi cuadrado frecuencias esperadas iguales

$$x^2 = \sum \left[\frac{(f_0 - f_e)^2}{f_e} \right]$$

Prueba de bondad de ajuste de frecuencias esperadas

EJEMPLO: Una empresa de venta de vehículos desea comprobar si no hay diferencia significativa en la venta de vehículos por sus vendedores, se espera que las frecuencias observadas (fo) fueran iguales. Puede concluirse que existe diferencia entre las ventas de vehículos de cada vendedor

Vendedor	**Vehículos**
A	13
B	33
C	14
D	7
E	36
F	17
TOTAL	**120**

Debido a que existen 120 datos, es de esperar que 20 queden en cada una de las 6 categorías.

Vendedores	**Vehículos vendidos fo**	**Número vendido esperado fe**
A	13	20
B	33	20
C	14	20
D	7	20
E	36	20
F	17	20

TOTAL	120	120

Paso 1. Se establece $H_0 \ y \ H_1$
$H_0 = \ f_0 = f_e$
$H_1 = \ f_0 \neq f_e$

Paso 2. Se selecciona el nivel de significancia 0.05, que es la probabilidad de rechazar una hipótesis nula verdadera

Paso 3. Selección del estadístico de prueba

$$x^2 = \sum \left[\frac{(f_0 - f_e)^2}{f_e} \right]$$

El estadístico es chi cuadrado, con $K - 1$ grados de libertad, donde:
K=es el número de categorías
f_0=es una frecuencia observada en una categoría determinada
f_e=es una frecuencia esperada en una categoría determinada

Paso 4. Se formula la regla de decisión

Grados de libertad	**Área de la cola derecha**			
gl	**0.10**	**0.05**	**0.02**	**0.01**
1	2,706	3,841	5,412	6,635
2	4,605	5,991	7,824	9,21
3	6,251	7,815	9,837	11,345
4	7,779	9,488	11,668	13,277
5	9,236	11,07	13,388	15,086

N= 6-1=5gdl
Se rechaza Ho si el valor ji cuadrada que se obtuvo de los cálculos es mayor que 11,070.

JUGADOR	**fo**	**fe**	**(fo-fe)**	**$(fo-fe)^2$**	**$(fo-fe)^2/fe$**
RONALDO	13	20	-7	49	2,45
BEKAM	33	20	13	169	8,45
ADRIANO	14	20	-6	36	1,8
DEKO	7	20	-13	169	8,45
RONALDIÑO	36	20	16	256	12,8
SIDANE	17	20	-3	9	0,45
TOTAL	120	120	13	519	34.5

Paso 5. Decidir.
Como el resultado calculado 34.5 es mayor que el de la tabla 11.070, rechazamos la hipótesis de que las frecuencias son iguales, las ventas son diferentes.

Prueba de bondad de ajuste
Frecuencias esperadas diferentes
Una empresa quiere comparar si el comportamiento de los datos de ingresos a un hospital obtenidos a nivel local difiere de los obtenidos a nivel nacional

Estudio nacional

Número de veces admitidas	**Porcentaje del total**
1	40
2	20
3	14
4	10

5	8
6	6
7	2
	100

Estudio local

Número de veces admitidas	**Número de personas, fo**
1	165
2	7
3	50
4	44
5	32
6	20
7	82
	400

A simple vista, no podemos comparar entre porcentajes y número de personas

Número de veces admitidas	**Número de personas, Fo**	**Número esperado de admisiones, fe (1) = (2) x (3)**		
1	165	160	40	400
2	7	80	20	400
3	50	56	14	400
4	44	40	10	400

5	32	32	8	400
6	20	24	6	400
7	10	8	2	400
	328	400	100	

Deben ser iguales

Paso. 1.
Ho: No existe diferencia entre la situación local y la situación nacional
H1: Si existe diferencia entre las situaciones local y nacional
Paso 2. Se establece el nivel de significancia de 0.05%
Paso 3. El estadístico de prueba a utilizar es chi cuadrado
Paso 4. Se establece la regla de decisión

Número de veces admitidas	**Número de personas, fo**	**Fe**	**Fo-Fe**	**(Fo-Fe)^2**	**(Fo-Fe)^2/Fe**
1	165	160	5	25	0,156
2	7	80	-1	1	0,013
3	50	56	-6	36	0,643
4	44	40	4	16	0,400
5	32	32	0	0	0,000
6	20	24	-4	16	0,667
7	10	8	2	4	0,500
	328	400			Chi =68.96

Observando el valor de la tabla con 7-1 grados de libertad, obtenemos un valor de 12,59. es decir, si el valor calculado de chi-cuadrado es mayor al valor de la tabla, entonces rechazamos Ho caso contrario aceptamos.

Paso 5. Decidir.

Como el valor calculado es 68,96 se encuentra en la región de Rechazo, es decir Rechazo Ho.

Referencia:
Puga-Ramón, I. R. (2023). Estadística Inferencial (p. 1). IES.