



UNL • FACULTAD DE
INGENIERÍA Y CIENCIAS HÍDRICAS

ESTADÍSTICA 2023

TRABAJO PRÁCTICO FINAL

Carrera:

- Ingeniería Informática

Alumno:

- Bargas, Santiago

Profesores:

- Vanlesberg, Susana
- Taulamet, Juan Pablo
- Llop, Maria José

INDICE

1. Introducción.	2
2. Objetivos.....	3
3. Desarrollo.	4
3.1. Cantidad total de ingresantes.	5
3.2. Cantidad de ingresantes varones.....	13
3.3. Cantidad de ingresantes mujeres.....	20
4. Conclusión:	29

1. Introducción.

En el siguiente trabajo se pretende aplicar los conocimientos sobre estadística adquiridos durante el cursado de la asignatura a un tema en específico: la cantidad de ingresantes en la carrera Ingeniería en Informática en la FICH.

Se eligió este tema específicamente ya que, la importancia y relevancia de la informática en nuestra sociedad han aumentado exponencialmente en los últimos años. Esta disciplina ha cambiado prácticamente todos los aspectos de nuestra vida diaria, desde cómo nos comunicamos hasta cómo realizamos las cosas que hacemos todos los días. El aumento constante en el número de estudiantes que optan por la carrera de Ingeniería Informática es uno de los indicadores más claros de esta expansión.

2. Objetivos.

En este trabajo se desea obtener y mostrar información sobre la cantidad de ingresantes en Ingeniería en Informática en relación del año correspondiente, con un periodo de 22 años (2002-2023) utilizando las herramientas aprendidas en estadística.

El objetivo es analizar y comprender cómo ha evolucionado la cantidad de ingresantes en Ingeniería Informática a lo largo de los años

Mediante este análisis estadístico, se espera obtener una visión más completa y precisa de cómo ha evolucionado la carrera informática y cómo estos cambios se han traducido en un aumento en el número de ingresantes en Ingeniería Informática.

3. Desarrollo.

En la Figura 1 se observa los datos proporcionados por la Universidad Nacional del Litoral, donde se puede ver el año correspondiente, la cantidad de ingresantes varones, la cantidad de ingresantes mujeres y la cantidad de total de ingresantes.

Para comenzar se trabajará con la cantidad de ingresantes total, luego con la cantidad de varones y mujeres por separado.

Año	Varones	Mujeres	Total
2002	289	58	347
2003	228	44	272
2004	241	50	291
2005	221	45	266
2006	226	34	260
2007	249	38	287
2008	218	45	263
2009	235	26	261
2010	202	33	235
2011	170	24	194
2012	183	17	200
2013	156	26	182
2014	172	16	188
2015	170	17	187
2016	185	15	200
2017	185	25	210
2018	231	28	259
2019	188	32	220
2020	305	59	364
2021	316	79	395
2022	365	75	440
2023	490	98	588

Figura 1. Cantidad de ingresantes por año.

3.1. Cantidad total de ingresantes.

Se comienza analizando la muestra con un análisis exploratorio (Figura 2) a partir de estadística descriptiva.

Año	Total
2002	347
2003	272
2004	291
2005	266
2006	260
2007	287
2008	263
2009	261
2010	235
2011	194
2012	200
2013	182
2014	188
2015	187
2016	200
2017	210
2018	259
2019	220
2020	364
2021	395
2022	440
2023	588

	Total de ingresantes
Media	277,6818181818182
Error estándar	21,037730109359945
Mediana	260,5
Moda	200
Desviación estándar	98,67570085585376
Varianza de la muestra	9736,89393939394
Curtosis	3,6139992177757985
Desviación	1,7601645235041228
Rango	406
Mínimo	182
Máximo	588
Suma	6109
Cuenta	22

Figura 2. Análisis exploratorio.

Datos:

- Media muestral (\bar{x}): 277,7
- Cantidad de muestra: $n= 22$
- Desvío estándar (S'): 98,67
- Desvío estándar de la población (σ): desconocido
- $1-\alpha=95\%$
- $\alpha= 5\%$
- Grados de libertad: 21

En la Figura 3 se observa un gráfico de box plot.

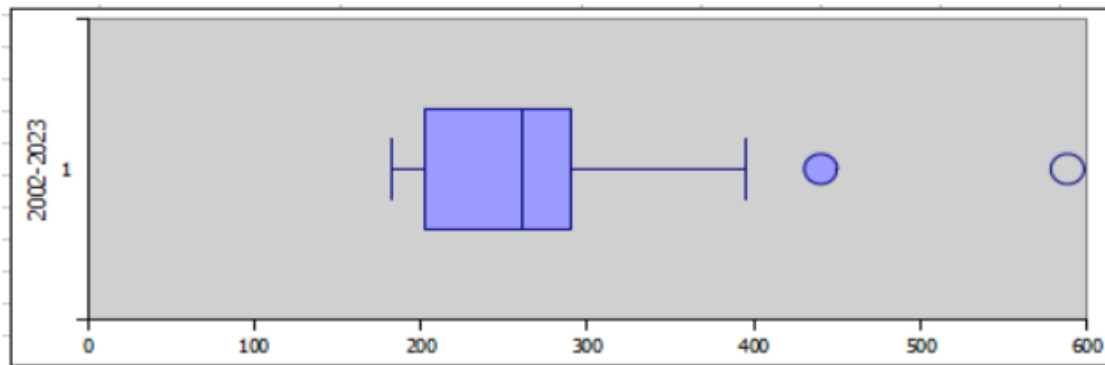


Figura 3. Gráfico de box plot.

En el gráfico se observa un valor atípico entre 400 y 500, que sería el correspondiente a la cantidad de ingresantes en 2022 y un valor atípico extremo en aproximadamente 600, que sería el correspondiente a la cantidad de ingresantes en 2023. Lo que se puede concluir a partir de este gráfico es que en los últimos dos años (2022 y 2023) hubo una cantidad atípica de ingresantes en la carrera Ingeniería en Informática en comparación de años anteriores.

Debido a que el tamaño de la caja es “chica” podemos decir que hay una varianza pequeña, donde la varianza es una medida de dispersión que indica qué tan dispersos están los valores de un conjunto de datos con respecto a su media. Y la línea del medio de la caja representa la mediana.

La mediana es un valor que divide un conjunto de datos en dos partes iguales: la mitad de los valores están por encima de la mediana y la otra mitad está por debajo. Se calcula ordenando los datos y seleccionando el valor central si el número de datos es impar, o el promedio de los dos valores centrales si el número de datos es par. La mediana es una medida de tendencia central que no se ve afectada por valores atípicos, lo que la hace útil para describir la ubicación central de un conjunto de datos.

Luego se puede estimar el intervalo de confianza en el que se encontrará el parámetro de población (μ). Para realizar este cálculo nos fijamos en la tabla de inferencia propuesta por la cátedra. Como tenemos que nuestra muestra es menor a 30 se obtendrá el intervalo de confianza por medio de la distribución normal, donde el desvío es parámetro desconocido, la fórmula que utilizamos es:

$$(xmedia + - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

El valor “t” lo calculamos como: Valor de t: = tinv(0,95)

$$(xmedia - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}}) \leq \mu \leq (xmedia + |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

Reemplazando los valores en la ecuación obtenemos un intervalo con un 95% de confianza de:

IC- 90% =	321,432172979	233,9314633847489
IC - 95% =	328,463638167	226,899998197
IC - 99% =	343,639441010	211,724195354

$$(226,9 \leq \mu \leq 328,5)$$

Este intervalo representa los rangos dentro de los cuales se espera que se encuentre la verdadera media poblacional de la cantidad de ingresantes en informática por año con un 95% de nivel de confianza. Cuanto mayor sea el nivel de confianza elegido, más amplio será el intervalo de confianza para tener una mayor probabilidad de capturar el verdadero valor poblacional. Sin embargo, esto también implica que el intervalo de confianza será más amplio y menos preciso en términos de la estimación puntual. Por lo tanto, la elección del nivel de confianza depende del equilibrio entre la precisión deseada y la probabilidad de capturar el verdadero valor.

La varianza es una medida de dispersión que indica qué tan dispersos están los datos alrededor de la media, como ya dijimos anteriormente.

Estimar la varianza poblacional proporciona información sobre la variabilidad de los valores en la población, lo que puede ayudar a comprender mejor la distribución de los datos y evaluar la dispersión de las observaciones.

Usando la tabla de inferencia, se ve que se puede estimar a través de la distribución chi-cuadrado.

Datos:

- $n=22$
- Desvío estándar (S')= 98,67
- $1 - \alpha = 95\%$

$$\chi^2 = \text{chiinv}(0,025; 21)$$

$$\chi^2_{inferior} = 35,47$$

$$\chi^2 = \text{chiinv}(0,95/2; 21)$$

$$\chi^2_{superior} = 20,73$$

Utilizamos la fórmula:

$$\frac{n * S^2}{\chi^2_{(1-\frac{\alpha}{2}; n-1)}} \leq \sigma^2 \leq \frac{n * S^2}{\chi^2_{(\frac{\alpha}{2}; n-1)}}$$

$$6037,72 \leq \sigma^2 \leq 10329,10$$

El hecho de que el intervalo sea relativamente amplio (desde 6037,72 hasta 10329,10) indica que existe una considerable variabilidad en la cantidad de ingresantes en informática por año en la población. Esto significa que los valores de ingresantes pueden diferir ampliamente de un año a otro.

También se realizó un histograma (Figura 4) con los datos y su respectivo polígono de frecuencias, donde un histograma es una representación gráfica que muestra la distribución de frecuencia de un conjunto de datos numéricos continuos. Consiste en un gráfico de barras donde el eje horizontal representa los rangos de valores y el eje vertical muestra la frecuencia o la densidad de los datos en cada rango.

Para determinar el número de cortes realizamos $cortes = \sqrt{n}$ donde n es la cantidad de muestras.

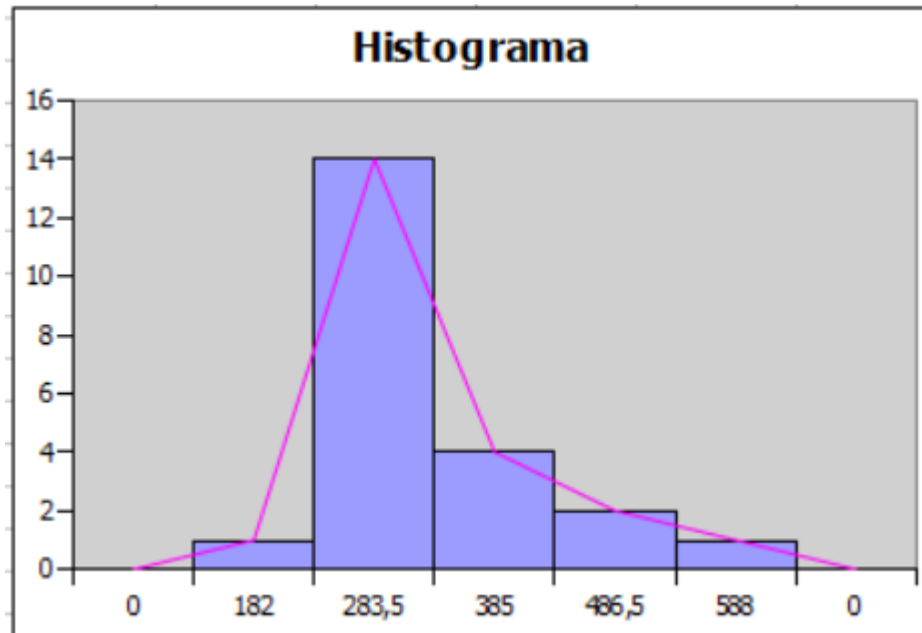


Figura 4. Histograma.

Siguiendo, el promedio móvil es una técnica de eliminar tanto como sea posible las variaciones indeseables de los datos, tratando de darle a la serie un aspecto más uniforme. El problema de esta técnica está en la elección apropiada del período para el promedio, esto depende de la naturaleza de los datos y el propósito que se persigue.

La suavización exponencial es un tipo especial de promedio móvil, pero su naturaleza es muy diferente del cálculo del Promedio Móvil.

$$S_i = \omega * Y_i + (1 - \omega) * S_{i-1}$$

S_i : Valor de la serie exponencialmente suavizada calculada en el período i

S_{i-1} Valor de la serie exponencialmente suavizada ya calculada en el período $i-1$

Y_i valor observado de la serie de tiempo en el período i

ω : ponderación o coeficiente de suavización asignado en forma subjetiva ($0 < \omega < 1$)

Sí solo se quiere suavizar una serie mediante la eliminación de variaciones cíclicas e irregulares que no se desean, debe seleccionarse un valor pequeño de ω (cercano a 0). Si se quieren hacer pronósticos se elegirá un valor grande de ω (cercano a 1). Para este caso si queremos hacer un pronóstico de la cantidad de ingresantes (Figura 5) para Ingeniería en Informática elegimos un ω cercano a uno como vemos a continuación.

Año	Total	X	W=0,25 0,25	W=0,5 0,5	W=0,75 0,75
2002	347	1	347,00	347,00	347,00
2003	272	2	328,25	309,50	290,75
2004	291	3	318,94	300,25	290,94
2005	266	4	305,70	283,13	272,23
2006	260	5	294,28	271,56	263,06
2007	287	6	292,46	279,28	281,01
2008	263	7	285,09	271,14	267,50
2009	261	8	279,07	266,07	262,63
2010	235	9	268,05	250,54	241,91
2011	194	10	249,54	222,27	205,98
2012	200	11	237,15	211,13	201,49
2013	182	12	223,37	196,57	186,87
2014	188	13	214,52	192,28	187,72
2015	187	14	207,64	189,64	187,18
2016	200	15	205,73	194,82	196,79
2017	210	16	206,80	202,41	206,70
2018	259	17	219,85	230,71	245,92
2019	220	18	219,89	225,35	226,48
2020	364	19	255,92	294,68	329,62
2021	395	20	290,69	344,84	378,66
2022	440	21	328,01	392,42	424,66
2023	588	22	393,01	490,21	547,17

Figura 5. Datos suavización exponencial.

Como se observa en la Figura 5, el último valor en $\omega = 0,75$ puede ser tomado como una aproximación para la cantidad de ingresantes en 2024 que es: 547,17.

Elegimos $\omega = 0,75$ pero no más cercano a 1, ya que al elegir un valor tan cercano a 1, se corre el riesgo de perder información importante sobre los cambios recientes en los datos. Esto puede hacer que el modelo sea menos capaz de capturar patrones o tendencias emergentes en tiempo real.

A continuación, en la Figura 6 se ve una gráfica con la serie de datos original y sus respectivas suavizaciones

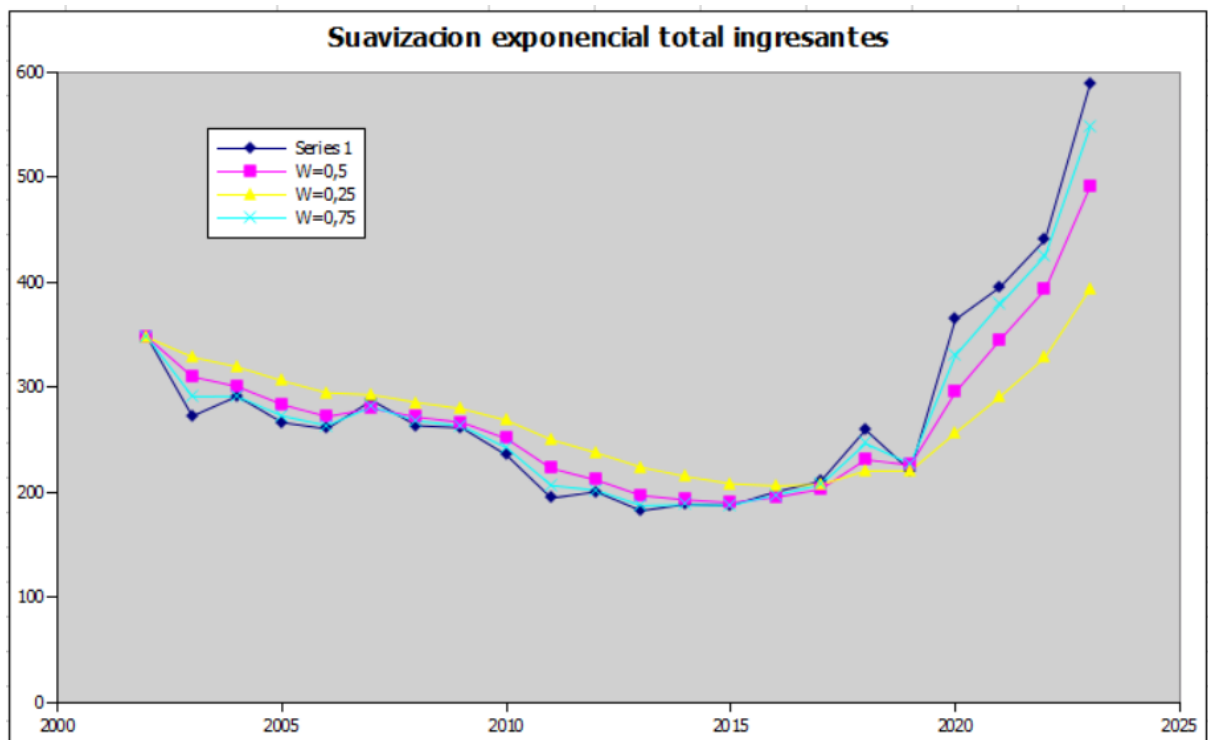


Figura 6. Suavización exponencial.

Luego, se realizó un dispersiograma (Figura 7) con los datos de la tabla tomando como variable X: “Año correspondiente” Y: “Cantidad de ingresantes en ingeniería informática”

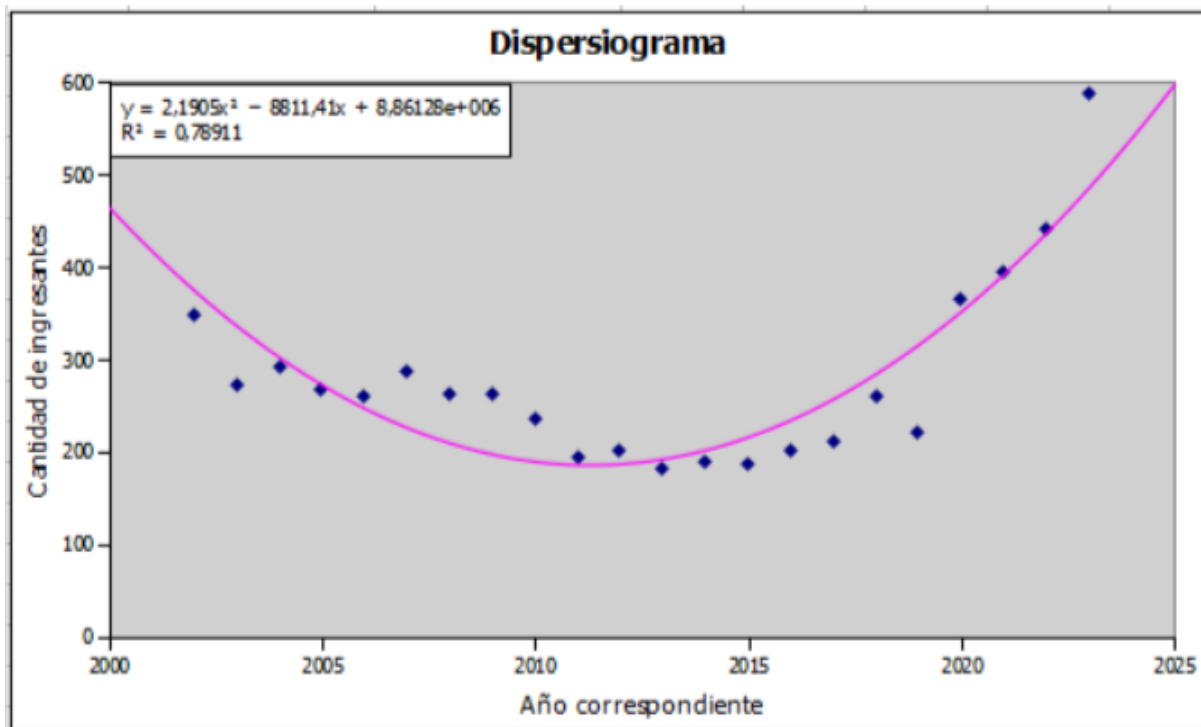


Figura 7. Dispersiograma cantidad total de ingresantes.

Como se observa en la Figura 7, el modelo que mejor se ajusta a los datos representados es un ajuste polinómico.

$$y = 2,1905 * x^2 - 8811,41 * x + 8861279,50$$

En la Figura 7 se ve que el coeficiente de determinación $r^2 = 0,78911$ es decir que aproximadamente el 78,9% de la variabilidad observada es explicada por la relación propuesta.

Como el modelo que mejor ajusta los datos es una ecuación polinómica, se procedió elevando al cuadrado los datos de la variable x, para así poder realizar una regresión polinómica. Se muestra a en la Figura 8 como se procedió

Año	Año ²	Total
2002	4008004	347
2003	4012009	272
2004	4016016	291
2005	4020025	266
2006	4024036	260
2007	4028049	287
2008	4032064	263
2009	4036081	261
2010	4040100	235
2011	4044121	194
2012	4048144	200
2013	4052169	182
2014	4056196	188
2015	4060225	187
2016	4064256	200
2017	4068289	210
2018	4072324	259
2019	4076361	220
2020	4080400	364
2021	4084441	395
2022	4088484	440
2023	4092529	588
2024		540,78571

Figura 8. Datos regresión.

En la hoja de regresión (Figura 9) podemos ver que $r = 0,888$ lo cual significa que hay una correlación positiva moderadamente fuerte entre las variables, lo que implica que a medida que aumenta una variable, la otra tiende a aumentar también.

Se puede destacar que la varianza de la regresión es: 2269,55. La varianza de la regresión, también conocida como varianza residual, es una medida de dispersión que indica cuánto varían los valores observados (y) con respecto a los valores predichos por la regresión lineal. Cuanto mayor es este valor, mayor es la variación de la curva ajustada respecto a las observaciones.

Para estimar la cantidad de ingresantes en el año 2024 reemplazamos en la ecuación $x = 2024$

$$y = 2,1905 * 2024^2 - 8811,41 * 2024 + 8861279,50$$

$$y = 540,78571$$

Lo cual nos dice que se estima que aproximadamente haya 541 ingresantes en total en 2024.

Es pertinente analizar r^2 ya que estamos en un caso que no es lineal. Para obtener ese $r^2 = 0,78$ se hace un cociente entre la suma de los cuadrados explicada por la regresión, y relacion en total.

SALIDA RESUMEN		Variable de respuesta		Columna 3			
Estadísticas de regresión							
R múltiple	0,8883188418019075						
R^2	0,7891103647002823						
Error estándar	47,639882809771905						
R^2 ajustado	0,7669114557213645					0,7891103647002823	
Observaciones	22						
Análisis de varianza							
	df	SS	MS	F		Significancia de F	
Regresión	2	161353,16247882554	80676,58123941277	35,547258896808934		3,78900117395597E-07	
Residual	19	43121,61024844721	2269,5584341288004				
Total	21	204474,77272727276					
	Coeficientes	Error estándar	Estadísticas-t	Valor-P		0,95	0,95
Interceptar	8861279,5	1146226,8840642981	7,730825042752114	2,77580167162069E-07		6462199,059843939	11260359,940156061
Columna 1	-8811,40676171654	1139,1154077102042	-7,73530645101958	2,75236471441148E-07		-11195,60271080113	-6427,210812631957
Columna 2	2,1904997176736307	0,2830097596777688	7,740014761920949	2,72796212063801E-07		1,5981534830357456	2,782845952311516

Figura 9. Hoja de regresión.

Ya explicado todo el procedimiento para la cantidad de ingresantes total por año, procedemos a realizar exactamente lo mismo, pero para la cantidad de ingresantes varones.

3.2. Cantidad de ingresantes varones.

Se comienza analizando la muestra con un análisis exploratorio (Figura 10) a partir de estadística descriptiva.

Año	Varones		
2002	289		
2003	228		
2004	241		
2005	221		
2006	226		
2007	249		
2008	218		
2009	235		
2010	202		
2011	170		
2012	183		
2013	156		
2014	172		
2015	170		
2016	185		
2017	185		
2018	231		
2019	188		
2020	305		
2021	316		
2022	365		
2023	490		

	Total de ingresantes
Media	237,5
Error estándar	16,559580217194423
Mediana	223,5
Moda	170
Desviación estándar	77,67131602678901
Varianza de la muestra	6032,833333333333
Curtosis	4,412122641685831
Desviación	1,9038703143332907
Rango	334
Mínimo	156
Máximo	490
Suma	5225
Cuenta	22

Figura 10. Análisis exploratorio.

Datos:

- Media muestral (\bar{x}): 237,5
- Cantidad de muestra: $n= 22$
- Desvío estándar (S'): 77,67
- Desvío estándar de la población (σ): desconocido
- $1-\alpha=95\%$
- $\alpha= 5\%$
- Grados de libertad: 21

En la Figura 11 se observa un gráfico de box plot.

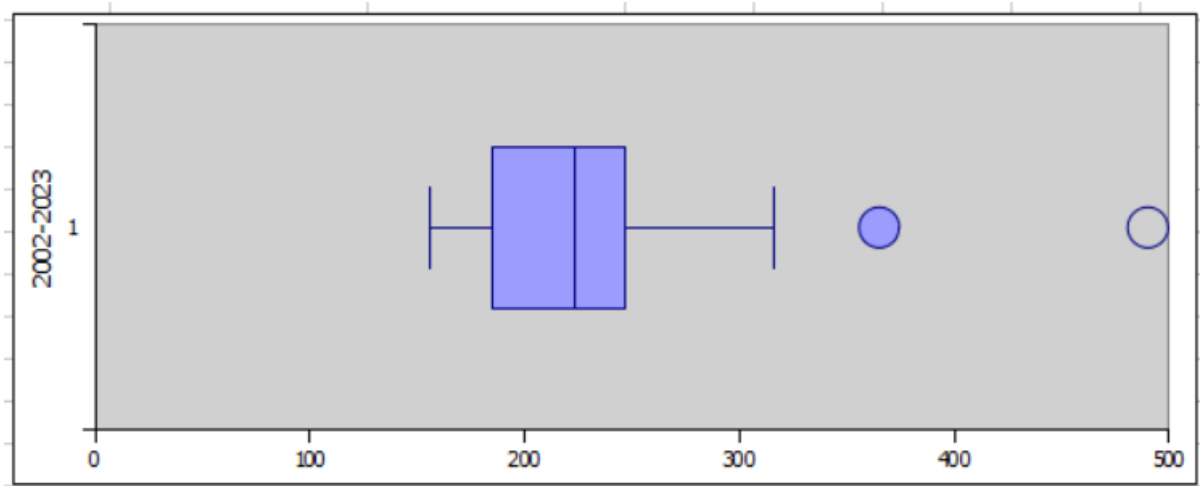


Figura 11. Gráfico de box plot.

Donde podemos ver que tiene un valor atípico en 400 aproximadamente que sería el correspondiente a la cantidad de ingresantes en 2022 y un valor atípico extremo en aproximadamente 500, que sería el correspondiente a la cantidad de ingresantes en 2023. Lo que se puede concluir a partir de este gráfico es que en los últimos dos años (2022 y 2023) hubo una cantidad atípica de ingresantes en la carrera Ingeniería en Informática en comparación de años anteriores.

Debido a que el tamaño de la caja es “chica” podemos decir que hay una varianza pequeña, donde la varianza es una medida de dispersión que indica qué tan dispersos están los valores de un conjunto de datos con respecto a su media. Y la línea del medio de la caja representa la mediana.

Estimamos el intervalo de confianza en el que se encontrará el parámetro de población (μ):

$$(xmedia + - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

El valor “t” lo calculamos como: Valor de t: = tinv(0,95)

$$(xmedia - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}}) \leq \mu \leq (xmedia + |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

Reemplazando los valores en la ecuación obtenemos un intervalo con un 95% de confianza de:

IC- 90% =	271,93753228256	203,062467717444
IC - 95% =	277,472260184	197,527739816
IC - 99% =	289,417699318	185,582300682

$$(197,52 \leq \mu \leq 277,57)$$

Este intervalo representa los rangos dentro de los cuales se espera que se encuentre la verdadera media poblacional de la cantidad de ingresantes varones en informática por año con un 95% de nivel de confianza.

Estimamos la varianza poblacional:

Datos:

- $n=22$
- Desvío estándar (S')= 98,67
- $1 - \alpha = 95\%$

$$\chi^2 = \text{chiinv}(0,025; 21)$$

$$\chi^2_{inferior} = 35,47$$

$$\chi^2 = \text{chiinv}(0,95/2; 21)$$

$$\chi^2_{superior} = 20,73$$

Utilizamos la fórmula:

$$\frac{n * S^2}{\chi^2_{(1-\frac{\alpha}{2}; n-1)}} \leq \sigma^2 \leq \frac{n * S^2}{\chi^2_{(\frac{\alpha}{2}; n-1)}}$$

$$3740,88 \leq \sigma^2 \leq 6399,75$$

Se realizó un histograma (Figura 12) con los datos y su respectivo polígono de frecuencias

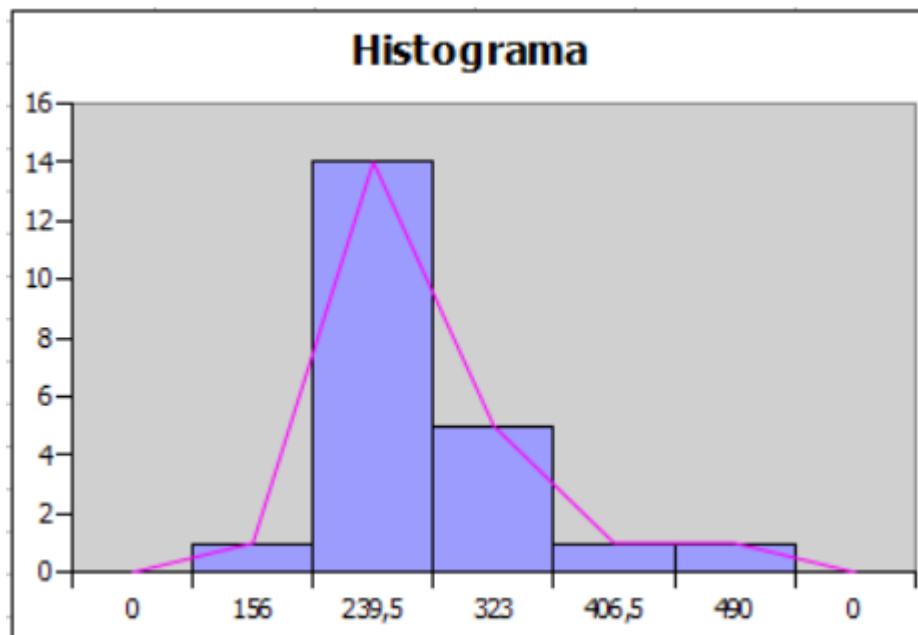


Figura 12. Histograma.

Siguiendo, con los respectivos datos se suavizó la serie (Figura 13).

			W=0,25	W=0,5	W=0,75
Año	Total	X	0,25	0,5	0,75
2002	289	1	289,00	289,00	289,00
2003	228	2	273,75	258,50	243,25
2004	241	3	265,56	249,75	241,56
2005	221	4	254,42	235,38	226,14
2006	226	5	247,32	230,69	226,04
2007	249	6	247,74	239,84	243,26
2008	218	7	240,30	228,92	224,31
2009	235	8	238,98	231,96	232,33
2010	202	9	229,73	216,98	209,58
2011	170	10	214,80	193,49	179,90
2012	183	11	206,85	188,25	182,22
2013	156	12	194,14	172,12	162,56
2014	172	13	188,60	172,06	169,64
2015	170	14	183,95	171,03	169,91
2016	185	15	184,21	178,02	181,23
2017	185	16	184,41	181,51	184,06
2018	231	17	196,06	206,25	219,26
2019	188	18	194,04	197,13	195,82
2020	305	19	221,78	251,06	277,70
2021	316	20	245,34	283,53	306,43
2022	365	21	275,25	324,27	350,36
2023	490	22	328,94	407,13	455,09

Figura 13. Datos suavización exponencial.

Como se observa en la Figura 13, el último valor en $\omega = 0,75$ puede ser tomado como una aproximación para la cantidad de ingresantes varones en 2024 que es: 455,09.

A continuación, en la Figura 14 vemos una gráfica con la serie de datos original y sus respectivas suavizaciones:

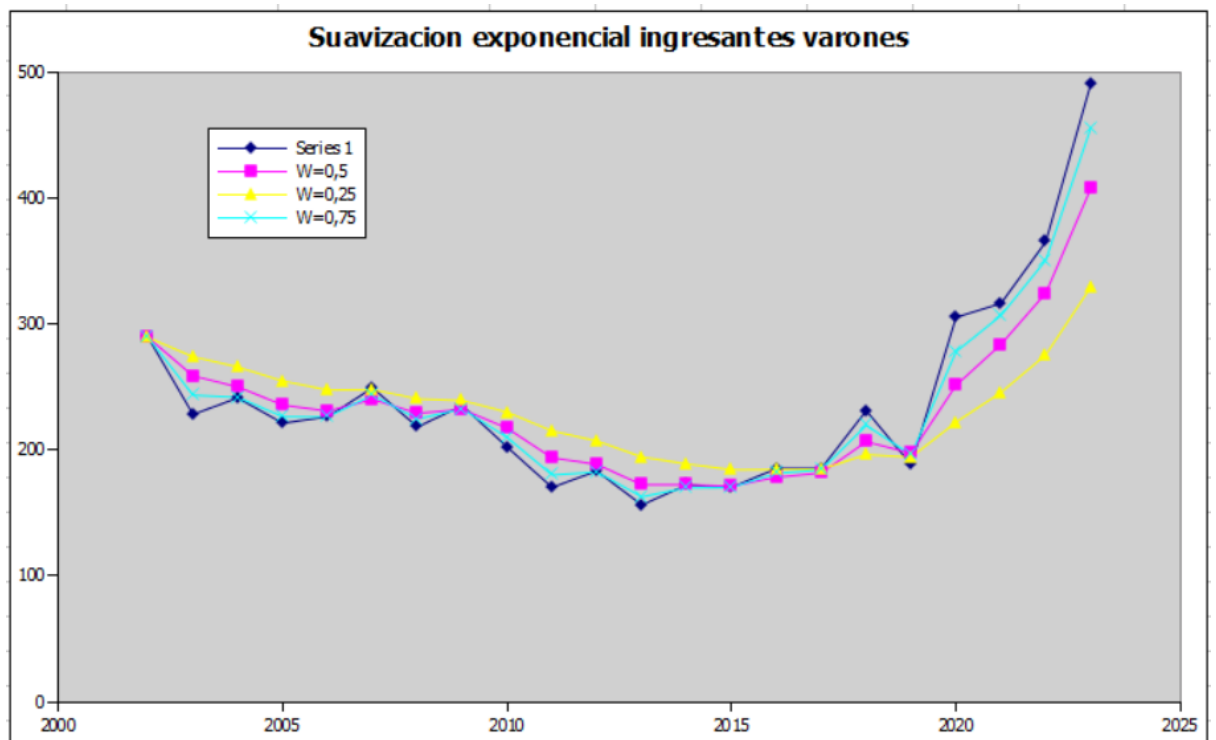


Figura 14. Suavización exponencial.

Luego, se realizó un dispersiograma (Figura 15) con los datos de la tabla tomando como variable X: “Año correspondiente” Y: “Cantidad de ingresantes varones en ingeniería informática”

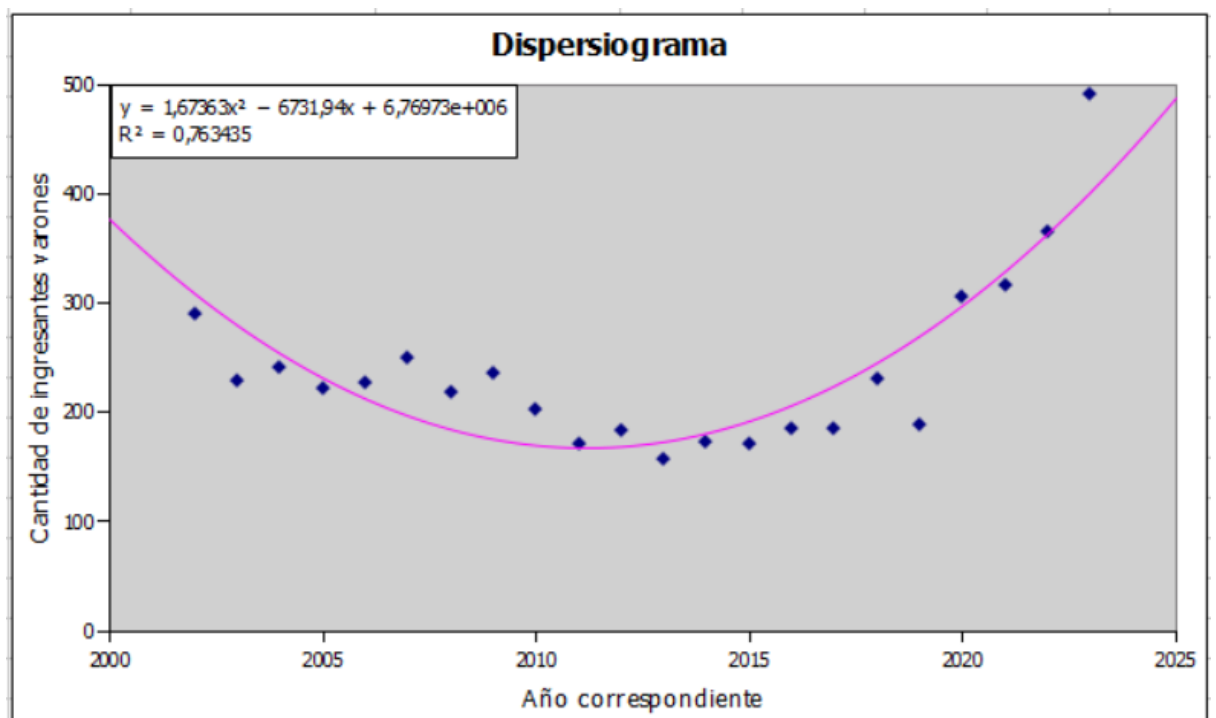


Figura 15. Dispersiograma

Como se observa en la Figura 15, el modelo que mejor se ajusta a los datos representados es un ajuste polinómico.

$$y = 1,67363 * x^2 - 6731,93732 * x + 6769728,09091$$

En la Figura 15 se ve que el coeficiente de determinación $r^2 = 0,763435$. Es decir que aproximadamente el 76,3% de la variabilidad observada es explicada por la relación propuesta.

Se realizó una regresión de los datos (Figura 16), exactamente igual como lo hicimos anteriormente.

Año	Año ²	Total
2002	4008004	289
2003	4012009	228
2004	4016016	241
2005	4020025	221
2006	4024036	226
2007	4028049	249
2008	4032064	218
2009	4036081	235
2010	4040100	202
2011	4044121	170
2012	4048144	183
2013	4052169	156
2014	4056196	172
2015	4060225	170
2016	4064256	185
2017	4068289	185
2018	4072324	231
2019	4076361	188
2020	4080400	305
2021	4084441	316
2022	4088484	365
2023	4092529	490
2024		442,37662

Figura 16. Datos regresión.

En la hoja de regresión (Figura 17) podemos ver que $r = 0,873$ lo cual significa que hay una correlación positiva moderadamente fuerte entre las variables.

Se puede destacar que la varianza de la regresión es: 1577,38

Para estimar la cantidad de ingresantes en el año 2024 reemplazamos en la ecuación $x = 2024$

$$y = 1,67363 * 2024^2 - 6731,93732 * 2024 + 6769728,09091$$

$$y = 442,38$$

Lo cual nos dice que se estima que aproximadamente haya 442 ingresantes varones en 2024.

SALIDA RESUMEN		Variable de respuesta		Columna 3		
Estadísticas de regresión						
R múltiple	0,8737476652392238					
R^2	0,7634349825109946					
Error estándar	39,71629904259336					
R^2 ajustado	0,7385334017226782				0,7634349825109946	
Observaciones	22					
Análisis de varianza						
df		SS	MS	F	Significancia de F	
Regresión	2	96719,19621682665	48359,598108413324	30,658093114682618	1,12857380964967E-06	
Residual	19	29970,30378317335	1577,3844096407026			
Total	21	126689,5				
	Coefficientes	Error estándar	Estadísticas-t	Valor-P	0,95	0,95
Interceptar	6769728,090909091	955583,5785729469	7,084391405112772	9,70203791101299E-07	4769668,674958339	8769787,506859843
Columna 1	-6731,93732354602	949,654900649004	-7,08882491834175	9,61727674155282E-07	-8719,587873991119	-4744,286773100919
Columna 2	1,673630717108978	0,2359388727343631	7,093492893785549	9,52886060718438E-07	1,179804981105975	2,167456453111981

Figura 17. Hoja de regresión.

Ya realizado el procedimiento para la cantidad de ingresantes varones, procedemos a realizarlo para las mujeres.

3.3. Cantidad de ingresantes mujeres.

Se comienza analizando la muestra con un análisis exploratorio (Figura 18) a partir de estadística descriptiva.

Año	Mujeres			
2002	58			
2003	44			
2004	50			
2005	45			
2006	34			
2007	38			
2008	45			
2009	26			
2010	33			
2011	24			
2012	17			
2013	26			
2014	16			
2015	17			
2016	15			
2017	25			
2018	28			
2019	32			
2020	59			
2021	79			
2022	75			
2023	98			
		Total de ingresantes		
		Media	40,181818181818	
		Error estándar	4,7421095839073	
		Mediana	33,5	
		Moda	45	
		Desviación estándar	22,242465527168	
		Varianza de la muestra	494,72727272727	
		Curtosis	0,8467484965356	
		Desviación	1,123857858023	
		Rango	83	
		Mínimo	15	
		Máximo	98	
		Suma	884	
		Cuenta	22	

Figura 18. Análisis exploratorio.

Datos:

- Media muestral (\bar{x}): 40,18
- Cantidad de muestra: $n= 22$
- Desvío estándar (S'): 22,24
- Desvío estándar de la población (σ): desconocido
- $1-\alpha=95\%$
- $\alpha= 5\%$
- Grados de libertad: 21

A continuación, se realizó un gráfico de box plot (Figura 19).

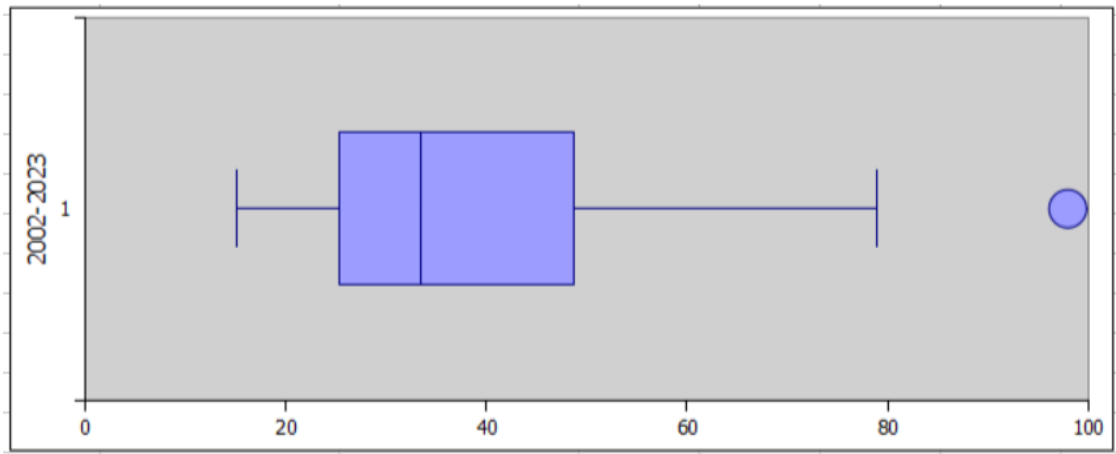


Figura 19. Gráfico de box plot.

Donde podemos ver que tiene un valor atípico aproximadamente en 100, que sería el correspondiente a la cantidad de ingresantes mujeres en 2023.

Estimamos el intervalo de confianza en el que se encontrará el parámetro de población (μ):

$$(xmedia + - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

El valor “t” lo calculamos como: Valor de t: = tinv(0,95)

$$(xmedia - |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}}) \leq \mu \leq (xmedia + |t_{(1-\frac{\alpha}{2}; n-1)}| \frac{S'}{\sqrt{n}})$$

Reemplazando los valores en la ecuación obtenemos un intervalo con un 95% de confianza de:

IC- 90% =	50,043574925728	30,3200614379088
IC - 95% =	51,628535769	28,735100595
IC - 99% =	55,049309766	25,314326598

$$(28,735 \leq \mu \leq 51,62)$$

Estimamos la varianza poblacional:

Datos:

- n=22
- Desvío estándar (S')= 22,24
- $1 - \alpha = 95\%$

$$\chi^2 = chiinv(0,025; 21)$$

$$\chi^2_{inferior} = 35,47$$

$$\chi^2 = chiinv(0,95/2; 21)$$

$$\chi^2_{superior} = 20,73$$

Utilizamos la fórmula:

$$\frac{n * S^2}{\chi^2_{(1-\frac{\alpha}{2}; n-1)}} \leq \sigma^2 \leq \frac{n * S^2}{\chi^2_{(\frac{\alpha}{2}; n-1)}}$$

$$306,77 \leq \sigma^2 \leq 524,82$$

Se realizó un histograma (Figura 20) con los datos y su respectivo polígono de frecuencias

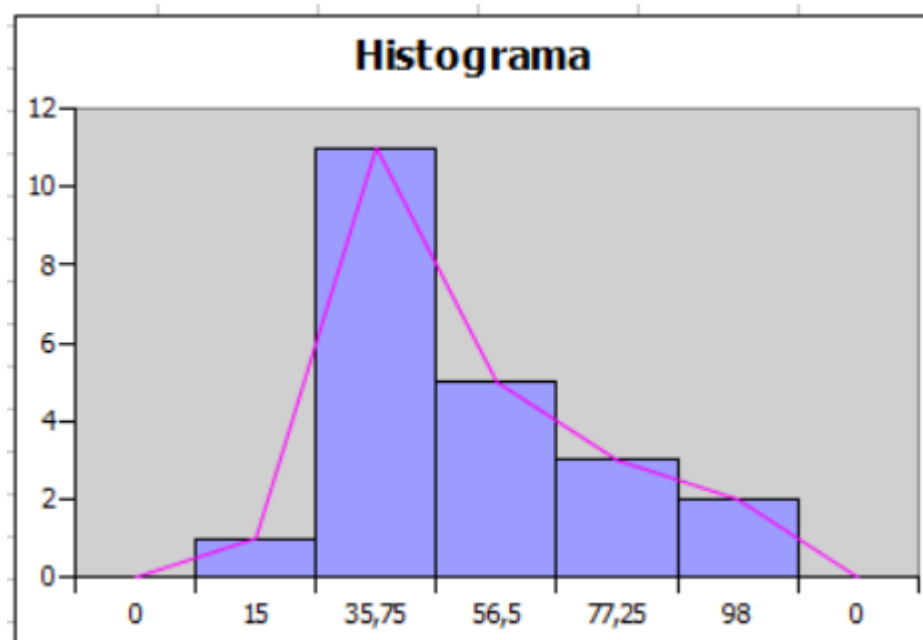


Figura 20. *Histograma.*

Siguiendo, con los respectivos datos se suavizó la serie (Figura 21).

Año	Total	X	W=0,25 0,25	W=0,5 0,5	W=0,75 0,75
2002	58	1	58,00	58,00	58,00
2003	44	2	54,50	51,00	47,50
2004	50	3	53,38	50,50	49,38
2005	45	4	51,28	47,75	46,09
2006	34	5	46,96	40,88	37,02
2007	38	6	44,72	39,44	37,76
2008	45	7	44,79	42,22	43,19
2009	26	8	40,09	34,11	30,30
2010	33	9	38,32	33,55	32,32
2011	24	10	34,74	28,78	26,08
2012	17	11	30,30	22,89	19,27
2013	26	12	29,23	24,44	24,32
2014	16	13	25,92	20,22	18,08
2015	17	14	23,69	18,61	17,27
2016	15	15	21,52	16,81	15,57
2017	25	16	22,39	20,90	22,64
2018	28	17	23,79	24,45	26,66
2019	32	18	25,84	28,23	30,67
2020	59	19	34,13	43,61	51,92
2021	79	20	45,35	61,31	72,23
2022	75	21	52,76	68,15	74,31
2023	98	22	64,07	83,08	92,08

Figura 21. Datos suavización exponencial.

Como se observa en la Figura 21, el último valor en $\omega = 0,75$ puede ser tomado como una aproximación para la cantidad de ingresantes varones en 2024 que es: 92,08

A continuación, en la Figura 22 vemos una gráfica con la serie de datos original y sus respectivas suavizaciones:

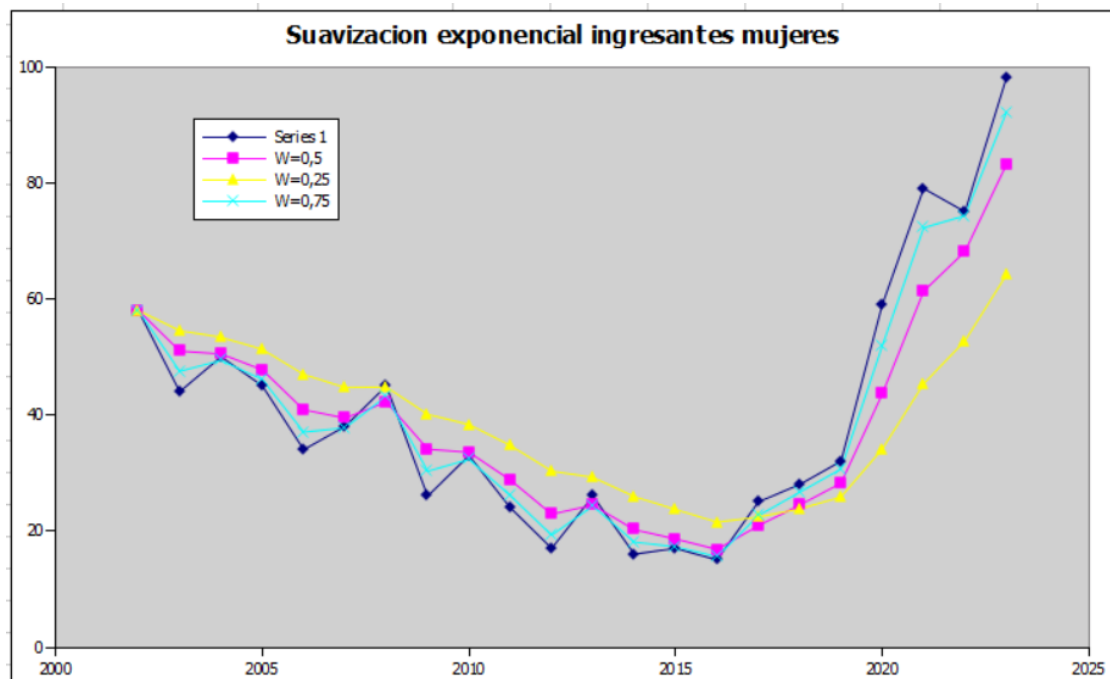


Figura 22. Suavización exponencial.

Luego, se realizó un dispersiograma (Figura 23) con los datos de la tabla tomando como variable X: “Año correspondiente” Y: “Cantidad de ingresantes mujeres en ingeniería informática”

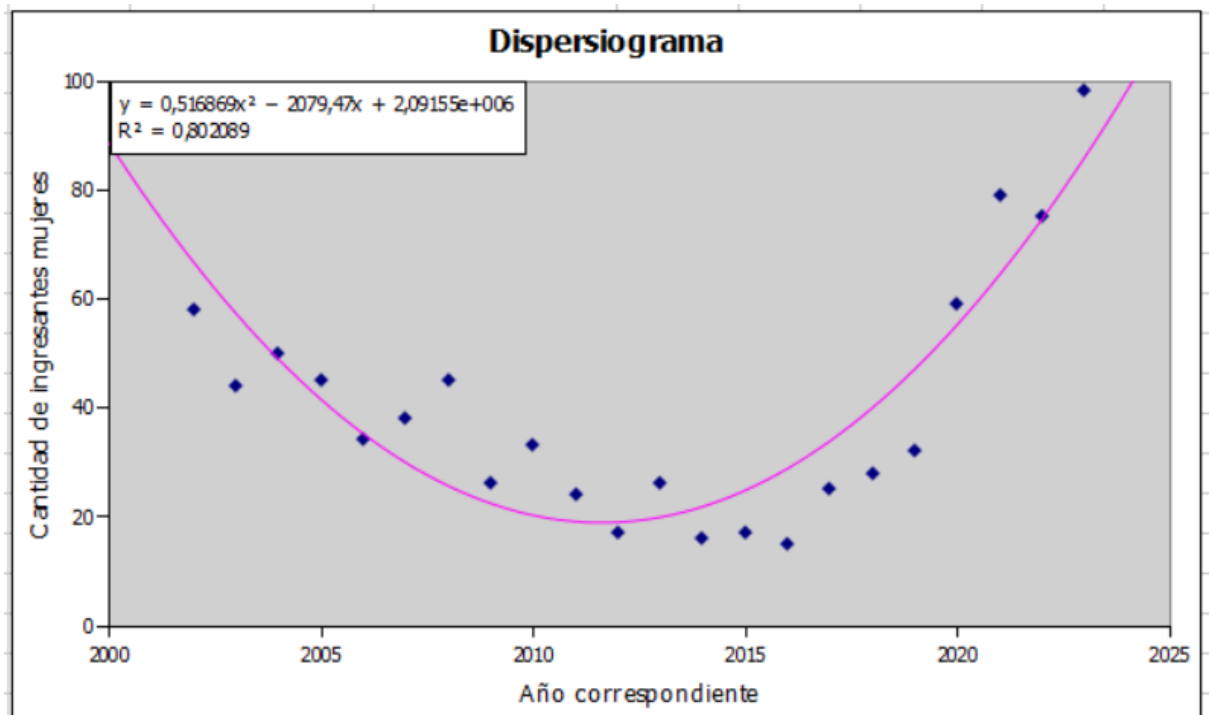


Figura 23. Dispersiograma

Como se observa en la Figura 23, el modelo que mejor se ajusta a los datos representados es un ajuste polinómico.

$$y = 0,51687 * x^2 - 2079,46944 * x + 2091551,40909$$

En la Figura 23 se ve que el coeficiente de determinación $r^2 = 0,802$. Es decir que aproximadamente el 80,2% de la variabilidad observada es explicada por la relación propuesta.

Se realizó una regresión de los datos (Figura 24), exactamente igual como lo hicimos anteriormente.

Año	Año ²	Total
2002	4008004	58
2003	4012009	44
2004	4016016	50
2005	4020025	45
2006	4024036	34
2007	4028049	38
2008	4032064	45
2009	4036081	26
2010	4040100	33
2011	4044121	24
2012	4048144	17
2013	4052169	26
2014	4056196	16
2015	4060225	17
2016	4064256	15
2017	4068289	25
2018	4072324	28
2019	4076361	32
2020	4080400	59
2021	4084441	79
2022	4088484	75
2023	4092529	98
2024		98,40909

Figura 24. Datos regresión.

En la hoja de regresión (Figura 25) podemos ver que $r = 0,895$ lo cual significa que hay una correlación positiva moderadamente fuerte entre las variables.

Se puede destacar que la varianza de la regresión es: 108,21

Para estimar la cantidad de ingresantes en el año 2024 reemplazamos en la ecuación $x = 2024$

$$y = 0,51687 * 2024^2 - 2079,46944 * 2024 + 2091551,40909$$

$$y = 98,409$$

Lo cual nos dice que se estima que aproximadamente haya 98 ingresantes mujeres en 2024.

SALIDA RESUMEN		Variable de respuesta		Columna 3				
Estadísticas de regresión								
R múltiple	0,8955939824864022							
R^2	0,8020885814658542							
Error estándar	10,402822752029373							
R^2 ajustado	0,781255800567523					0,8020885814658543		
Observaciones	22							
Análisis de varianza								
	df	SS	MS	F		Significancia de F		
Regresión	2	8333,117024280069	4166,5585121400345	38,5012728439969		2,07239764665761E-07		
Residual	19	2056,1557029926594	108,21872121013996					
Total	21	10389,272727272728						
Coeficientes								
	Error estándar	Estadísticas-t	Valor-P		0,95			0,95
Interceptar	2091551,4090909092	250294,38372350406	8,356365724136305	8,72037413255682E-08	1567679,2432743113	2615423,574907507		
Columna 1	-2079,46943817053	248,74149518445603	-8,35996196223103	8,6638017943016E-08	-2600,091370921081	-1558,84750541997		
Columna 2	0,5168690005646528	0,0617990682046423	8,363702165428869	8,6053693055492E-08	0,3875220642723167	0,6462159368569889		

Figura 25. Hoja de regresión.

Ya finalizado el procedimiento por separado para la cantidad total de ingresantes, la cantidad de varones y la cantidad de mujeres, podemos decir que parece repetitivo el proceso e innecesario pero la verdad es que esto nos sirve para estimar por separado cada componente en los siguientes años, ya que obtuvimos sus ecuaciones que se ajustan al modelo independientemente de las otras.

En la Figura 26 se observa la aproximación de las componentes por separado para algunos años posteriores.

Año	Varones	Mujeres	Total
2002	289	58	347
2003	228	44	272
2004	241	50	291
2005	221	45	266
2006	226	34	260
2007	249	38	287
2008	218	45	263
2009	235	26	261
2010	202	33	235
2011	170	24	194
2012	183	17	200
2013	156	26	182
2014	172	16	188
2015	170	17	187
2016	185	15	200
2017	185	25	210
2018	231	28	259
2019	188	32	220
2020	305	59	364
2021	316	79	395
2022	365	75	440
2023	490	98	588
2024	442,38	98,41	540,79
2025	486,97	111,74	598,71
2026	534,91	126,11	661,02
2027	586,20	141,51	727,71
2028	640,83	157,94	798,78
2029	698,82	175,41	874,23
2030	760,15	193,91	954,06
2031	824,82	213,45	1038,27
2032	892,85	234,02	1126,87
2033	964,22	255,62	1219,84

Figura 26. Estimación de ingresantes años posteriores.

Donde para aproximar desde el año 2024 hasta el año 2033 se utilizaron sus ecuaciones correspondientes:

$$\text{Cant. total: } y = 2,1905 * x^2 - 8811,41 * x + 8861279,50$$

$$\text{Cant. varones: } y = 1,67363 * x^2 - 6731,93732 * x + 6769728,09091$$

$$\text{Cant. mujeres: } y = 0,51687 * x^2 - 2079,46944 * x + 2091551,40909$$

Para obtener el valor de y , se reemplazó x por el valor correspondiente al año.

Para terminar, podemos pensar en la oportunidad en la que un profesor quiere elegir un alumno al azar, en este caso estudiaremos la probabilidad de que ese alumno seleccionado al azar sea varón o mujer.

En el año 2023, según los datos proporcionados, hubo 490 ingresantes varones y 98 ingresantes mujeres y un total de 588.

Para calcular la probabilidad de que el alumno elegido al azar sea mujer, dividimos la cantidad de ingresantes mujeres (98) entre el total de ingresantes (588):

$$\text{Probabilidad de que sea mujer} = 98 / 588 \approx 0.1667$$

Por lo tanto, la probabilidad de que el alumno elegido al azar del año 2023 sea mujer es aproximadamente 0.1667 o 16.67%.

De manera similar, para calcular la probabilidad de que el alumno elegido al azar sea hombre, dividimos la cantidad de ingresantes varones (490) entre el total de ingresantes (588):

$$\text{Probabilidad de que sea hombre} = 490 / 588 \approx 0.8333$$

Por lo tanto, la probabilidad de que el alumno elegido al azar del año 2023 sea hombre es aproximadamente 0.8333 o 83.33%

En este caso, hemos utilizado la probabilidad a priori, donde la probabilidad a priori es:

Si un procedimiento tiene N eventos simples diferentes que son igualmente probables, y si el evento puede ocurrir de n diferentes maneras entonces:

$$P(A) = \frac{\text{casos favorables } n}{\text{casos igualmente posibles } N}$$

4. Conclusión.

En el campo de la informática, hemos sido testigos de una rápida evolución en los últimos años. La forma en que vivimos, trabajamos y nos relacionamos ha sido significativamente afectada por los avances tecnológicos y la creciente dependencia de la sociedad de la tecnología. La revolución digital ha despertado un gran interés en la informática, lo que ha llevado a un aumento en el número de personas que deciden estudiar esta carrera.

La informática ha dejado de ser vista simplemente como una herramienta y se ha convertido en una parte importante de todas las facetas de nuestras vidas. La necesidad de desarrollar y mantener sistemas y aplicaciones que fomenten la eficiencia, la productividad y la innovación en una variedad de industrias, como la salud, las finanzas y el entretenimiento, aumenta la demanda de profesionales capacitados en informática.

También, el auge de la inteligencia artificial, el análisis de datos, la ciberseguridad y el internet de las cosas ha generado nuevas oportunidades laborales en el campo de la informática. Estos avances tecnológicos han despertado el interés de muchas personas en seguir una carrera en informática, ya que perciben el potencial de contribuir a la resolución de desafíos complejos y a la transformación digital de las organizaciones. Hubiese sido interesante poder realizar este trabajo sobre la nueva carrera Ingeniería Artificial en la FICH. Lamentablemente es una carrera nueva en esta facultad, donde comenzó a dictarse en 2023 y por la poca cantidad de datos no iba a ser posible realizarlo.

Además, al realizar análisis exploratorios de los datos, como el uso de gráficos de dispersión, box plots, histogramas y técnicas como suavización exponencial, estimación de la varianza poblacional, se ha obtenido una comprensión más profunda de las tendencias y variabilidad en los datos de ingresantes a la carrera de Ingeniería Informática.

En resumen, la evolución acelerada de la informática en los últimos años, impulsada por avances tecnológicos y una mayor dependencia de la tecnología, ha llevado a un creciente interés en estudiar informática. La demanda de profesionales en este campo ha aumentado debido a la necesidad de desarrollar soluciones innovadoras y eficientes en diferentes sectores. Además, la accesibilidad a recursos de aprendizaje y programas educativos ha permitido que más personas puedan perseguir una carrera en informática y contribuir al avance tecnológico en nuestra sociedad.