



UNIVERSIDAD DE INVESTIGACIÓN Y TECNOLOGÍA EXPERIMENTAL YACHAY TECH

Escuela de Ciencias Matemáticas y Computacionales

TÍTULO: Trabajo práctico de Probabilidad y Estadística

Autores:

Asero Oliver	#1751544691	TB
Murillo Dayana	#0705136307	TD
Pincay Ariel	#0942124652	TD
Tipan Joseph	#0954811246	TD

Tutor:

Isidro Amaro, PhD.

Urcuquí, marzo 2023

Resumen

La probabilidad es una herramienta matemática útil para entender y cuantificar la incertidumbre en diferentes situaciones y para hacer predicciones basadas en datos estadísticos. En el presente trabajo se aplicó algunos temas de probabilidad como fueron la elaboración de tabla de frecuencias, histogramas, medidas de tendencia central y dispersión, probabilidad de distribución normal, entre otros con ayuda del lenguaje de programación Python. Con el propósito de analizar los datos obtenidos y poder determinar la probabilidad de que suceda una determinada situación.

Palabras claves: *estadística, promedio, desviación estándar, varianza, histograma, distribución normal.*

Trabajo práctico de probabilidad y estadística

Los datos que se analizarán en el siguiente trabajo son los siguientes:

42,65	35,92	39,26	39,19	38,99
34,4	37,86	38,84	54,12	38,47
40,35	48,62	35,95	36,17	29,62
37,17	37,87	34,51	50,43	39,52
39,57	38,93	37,17	29,75	36,03
33,03	35,82	29,65	39,7	35,68
31,82	31,71	29,42	31,12	30,64
20,44	37,91	48,12	40,9	24,35
30,07	34,76	36,71	28,56	32,07
40,45	36,54	30,71	37,37	31,79
36,53	28,65	25,38	40,81	24,56
38,92	37,92	35,29	27,95	42,09
34,73	51,8	42,46	33,62	33,41
37,74	37,98	36,83	45,35	38,85
40,8	37,98	43,8	42,91	28,56
33,77	31,53	39,2	35,66	30,58
34,3	42,59	42,93	34,12	29,11
49,04	28,91	39,15	33,88	34,37
41,27	46,11	37,47	35,54	38,17
43,58	36,58	47,21	32,86	39,58

Con dichos datos, en el presente proyecto se buscó en primer lugar realizar una tabla de frecuencias, para lo cual se utilizó el método de Sturges, cual es una importante técnica comúnmente utilizada para determinar el número de clases adecuado en un histograma el que se lo realizara brevemente. El número de clases en un histograma se refiere al número de intervalos que se utilizan para dividir los datos en el eje horizontal. Este método se basa en la suposición de que la distribución de los datos es aproximadamente normal y no debería de tener una distribución sesgada o anormal, la ecuación de Sturges es la siguiente:

$$K = 1 + \log_2(n) \quad (1)$$

Donde " K " es el número de clases y " n " es el tamaño de la muestra.

N° de datos	100
Ley de Sturges	8
Valor Mínimo	20,44
Valor Máximo	54,12
Rango	33,68
Amplitud	4,40

Tabla 1. Valores necesarios para construir tabla de frecuencia

Una vez que se ha calculado el número de clases adecuado utilizando el método de Sturges, se puede construir una tabla de frecuencias para visualizar la distribución de los datos, la cual es una herramienta muy importante en la probabilidad y estadística; ya que, estas son capaces de permitir organizar y resumir los datos de la muestra.

Intervalo	Marca de clase	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia relativa	Frecuencia relativa acumulada	% de frecuencia	Frecuencia Porcentual Acumulada (%)
[20,40-24,80)	22,60	3	3	0,03	0,03	3	3
[24,80-29,20)	27,00	7	10	0,07	0,1	7	10
[29,20-33,60)	31,40	17	27	0,17	0,27	17	27
[33,60-38,00)	35,80	35	62	0,35	0,62	35	62
[38,00-42,40)	40,20	22	84	0,22	0,84	22	84
[42,40-46,80)	44,60	9	93	0,09	0,93	9	93
[46,80-51,20)	49,00	5	98	0,05	0,98	5	98
[51,20-55,60)	53,40	2	100	0,02	1	2	100
TOTAL		100		1		100	

Tabla 2. Tabla de frecuencias

A continuación, se logra visualizar la distribución utilizando un histograma de frecuencia y un histograma de frecuencia acumulada. Ambos tipos de gráficas se inspiran en la misma tabla de frecuencia, pero ambas proporcionan una información diferente de la distribución estudiada.

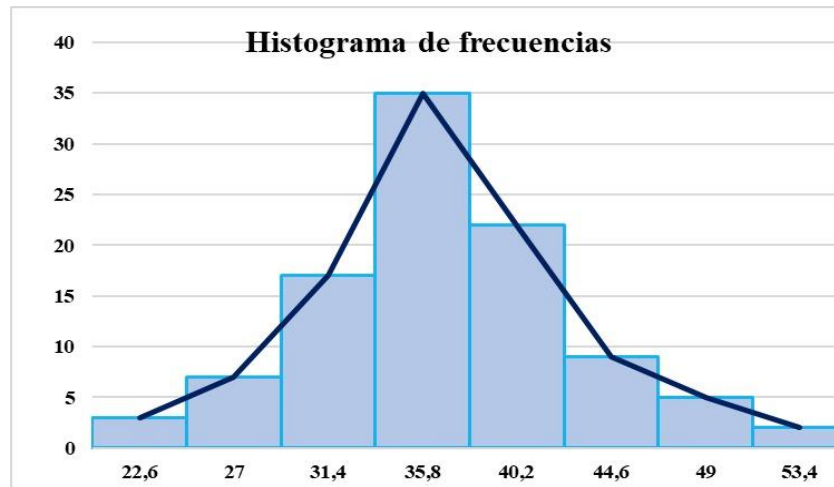


Figura 1. Histograma de frecuencias

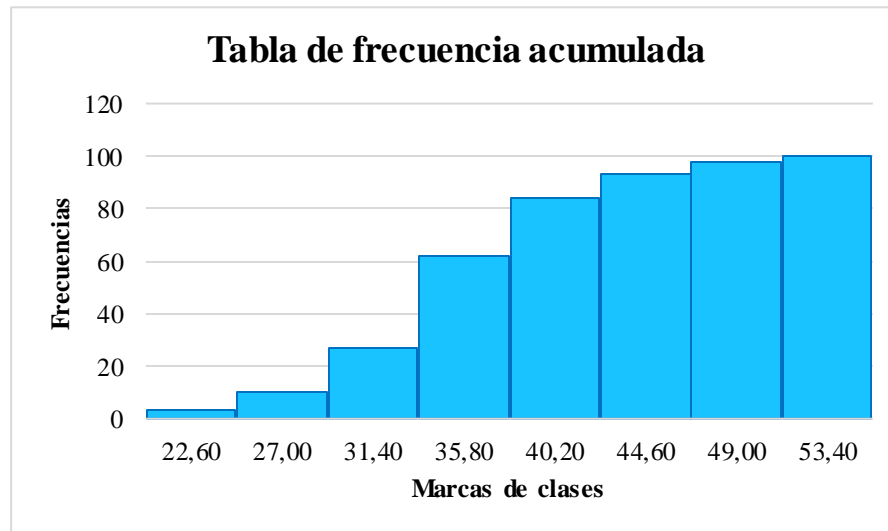


Figura 2. Histograma de frecuencia acumulada

Posteriormente se procede a calcular determinadas medidas de tendencia central y de dispersión, como fue la media y la varianza de los datos originales y los datos agrupados.

De muestra original (datos sin agrupar)		De datos agrupados	
Promedio	36,73	Promedio	36,81
Varianza	36,26	Varianza	39,45

Tabla 3. Promedio y varianza de los datos agrupados y sin agrupar

A partir de la tabla de frecuencias se buscó la probabilidad aproximada de que un objeto tomado al azar de esta población esté entre su media más o menos dos veces su desviación estándar.

Para la cual se aplica la probabilidad de distribución normal, primero sabemos que los valores

$$\mathbf{Var = \sigma^2 = 39,45}$$

$$\mathbf{Desviación\ estándar = \sigma = \sqrt{39,45} = 6,28}$$

Para encontrar los valores de la probabilidad de $P(24,25 \leq Y \leq 49,37)$ se necesita normalizar, para lo cual se utiliza la siguiente ecuación.

$$Z = \frac{Y - \mu}{\sigma} \quad (2)$$

El cálculo de dicha probabilidad se realizó mediante Phyton con el código [1], obteniendo como resultado una probabilidad de:

$$\mathbf{P(24,25 \leq Y \leq 49,37) = P(-2 \leq z \leq 2) = 0,9544}$$

Suponiendo que la variable aleatoria Y tiene una distribución como la de los datos con los parámetros pertinentes como los calculados, a continuación, se calculó las siguientes probabilidades:

$$\mathbf{P(Y < 9) = 0,24}$$

$$\mathbf{P(6 < Y < 9) = 0,023}$$

Además, se buscó el valor de k tal que $\mathbf{P(Y < k) = 0,4}$

Donde k es 26,82.

Por último, se encontró k, tal que $\mathbf{P(Y < k) = 0,4}$. Hallando que k es 4,02.

Anexos

Códigos usados

```
from scipy.stats import norm

# Definir la probabilidad acumulada
p = 0.3

# Calcular el valor de k
k = norm.ppf(p, loc=36.81, scale=39.45)

print("El valor de k es:", k)
```

```
[6] from scipy.stats import norm

media = 36.81
desviacion_estandar = 39.45

lim_inferior= 6
lim_superior= 9

probabilidad_inf = norm.cdf(lim_superior, loc=media, scale=desviacion_estandar)
probabilidad_sup = norm.cdf(lim_inferior, loc=media, scale=desviacion_estandar)

probabilidad = probabilidad_inf - probabilidad_sup

#valor = 8
#probabilidad = norm.cdf(valor, loc=media, scale=desviacion_estandar)

print("La probabilidad de que un valor esté por debajo de 2 desviaciones estándar de la media es:", probabilidad)
```

```
[7] from scipy.stats import norm

media = 36.81
desviacion_estandar = 39.45

valor = 8
probabilidad = norm.cdf(valor, loc=media, scale=desviacion_estandar)

print("La probabilidad de que un valor esté por debajo de 2 desviaciones estándar de la media es:", probabilidad)
```