

Anexo_Herramientas de estadística

La calidad de un producto está determinada por una serie de características o propiedades estéticas, funcionales, físicas, de durabilidad, que si se ajustan a las expectativas del cliente se tendrá una muy buena percepción sobre su valor; es por eso que cobra importancia el análisis y medición de esos datos, pues permite obtener información pertinente sobre la calidad del producto y generar acciones de mejora al funcionamiento del proceso y tomar decisiones basados en el análisis de dichos datos.

Para realizar este proceso se hace necesario recurrir a herramientas estadísticas que posibiliten información pertinente para visualizar y medir la variabilidad, para tomar decisiones y soluciones a los problemas de rendimiento y resultados de las empresas, basados en hechos y datos mejorando así los planes de acción.

Matriz de distribución de frecuencias

También conocida como cuadro de control estadístico, es una herramienta en la tabulación de datos que ayuda a ordenar la información de forma tal, que facilite la visualización y comprensión de los datos para obtener como resultado final una adecuada toma de decisiones. La matriz de frecuencias ordena en columnas los valores recogidos para estructurar variables y cómo se presentan en un proceso para luego tabularlos y formar ideas que ayudan a controlar y liderar los recursos humanos, maquinaria, materia prima y dependiente de las variables que se analizan.

En estadística, la distribución de frecuencias es una herramienta de agrupación de datos, ayuda a la trazabilidad de la información desde su punto de origen hasta obtener porcentuales de cada uno de los valores, muestra participación y como objetivo final tomamos decisiones más acertadas.

Contenido de una matriz de distribución de frecuencias

Variables

N = Tamaño poblacional.

X_i = Característica cuantitativa, observada en cada unidad investigada.

n_i = Frecuencia absoluta, número de veces que se repite cada valor de la variable

N_i = Frecuencia absoluta acumulada. Este resultado se obtiene dividiendo cada frecuencia absoluta por el tamaño de la muestra

F_i = frecuencia relativa acumulada o porcentual

Tabla 1

Matriz de distribución de frecuencias

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa ($f_i = n_i / N$)	Frecuencia relativa acumulada ($F_i = N_i / N$)
1	7	7	0,06	0,06
2	19	26	0,15	0,21
3	25	51	0,20	0,41
4	12	63	0,10	0,50
5	23	86	0,18	0,69
6	15	101	0,12	0,81
7	8	109	0,06	0,87
8	16	125	0,13	1,00
Total	125		1	

Nota. SENA (2021)

Lo atractivo de esta herramienta es el uso diario en las empresas y cumple el objetivo de visualizar fácilmente el estado real de los números estadísticos en el proceso.

Veamos a continuación su aplicación en los siguientes ejemplos:

Ejemplo 1

Los siguientes datos muestran las preferencias alimentarias en comidas caseras. Una muestra poblacional de 30 personas. Se pide con la información recolectada construir una matriz de distribución de frecuencia.

TAMAÑO MUESTRAL				
Hamburguesas	Frijoles	Pollo	Espaguetis	Pollo
Sancocho	Frijoles	Frijoles	Sudado	Frijoles
Frijoles	Sancocho	Frijoles	Pollo	Frijoles
Espaguetis	Hamburguesas	Sancocho	Hamburguesas	Espaguetis
Sudado	Sancocho	Espaguetis	Sancocho	Sudado
Pollo	Frijoles	Frijoles	Frijoles	Pollo

Con base en la información recolectada se crea la matriz de distribución de frecuencia:

Matriz de distribución de frecuencias preferencias alimentarias

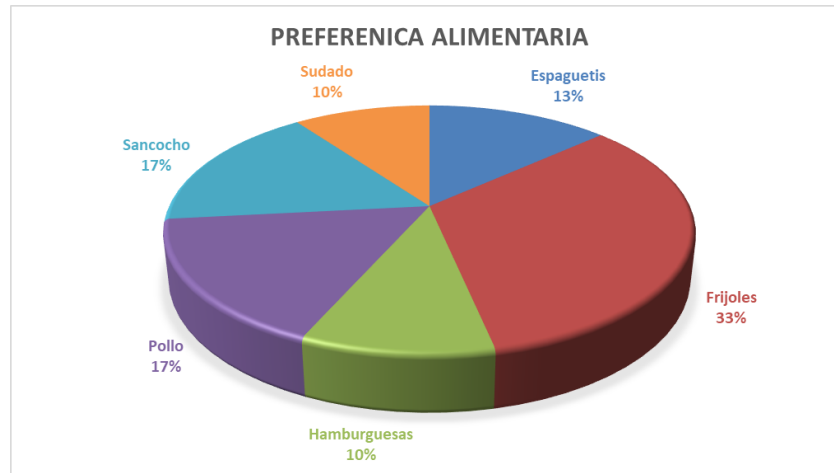
MATRIZ DE DISTRIBUCIÓN DE FRECUENCIA					
Variable	Frecuencia absoluta (n _i)	Frecuencia absoluta acumulada (N _i)	Frecuencia relativa (fi = n _i / N)	Frecuencia relativa acumulada (Fi = N _i / N)	Frecuencia porcentual
Espaguetis	4	4	0,13	0,13	13%
Frijoles	10	14	0,33	0,47	33%
Hamburguesas	3	17	0,10	0,57	10%
Pollo	5	22	0,17	0,73	17%
Sancocho	5	27	0,17	0,90	17%
Sudado	3	30	0,10	1,00	10%
Total	30		1		100%

Nota. SENA (2021)

Gráficamente así se observa la frecuencia porcentual de la preferencia alimentaria.

Figura 1

Preferencia alimentaria



Nota. SENA (2021)

Vemos entonces cómo el objetivo de la estadística es tomar parte activa para mostrar de forma simple la información. Validando, para la toma de decisiones se interpreta que el 33 % de la población tiene de plato predilecto los frijoles.

Ejemplo 2

Interpretemos con otro ejemplo de producción tomando como base los datos aportados por una planta de producción durante una semana de trabajo, se pide interpretar la información que se muestra en los siguientes cuadros.

Tabla 2

Registro muestral semanal

REGISTRO MUESTRAL DE LA PRODUCCIÓN POR SEMANA						
META HORARIA DE PRODUCCION UNIDADES						75
CONTROL HORARIO	LUNES	MARTES	MIERCOLES	JUEVES	VIERNES	SABADO
1 (6 - 7)	70	70	68	70	65	71
2 (7 -8)	71	71	71	71	71	80
3 (8 -9)	71	68	70	71	71	80
4 (9 -10)	70	68	70	59	72	70
5 (10 -11)	70	63	70	70	70	85
6 (11-12)	80	80	80	80	80	70
7 (12 -13)	67	76	76	76	76	
8 (13 -14)	77	77	77	77	77	
9 (14-15)	60	61	62	63	64	
10 (15 -16)	50	50	50	50	50	
PROMEDIO DIA	68,6	68,4	69,4	68,7	69,6	76
TOTAL	686	684	694	687	696	456
PORCENTUAL DE EFICIENCIA PROMEDIO DIARIA	91%	91%	93%	92%	93%	101%

Nota. SENA (2021)

Con base en la información recolectada se crea la matriz de distribución de frecuencia:

Tabla 3

Matriz de distribución de frecuencias

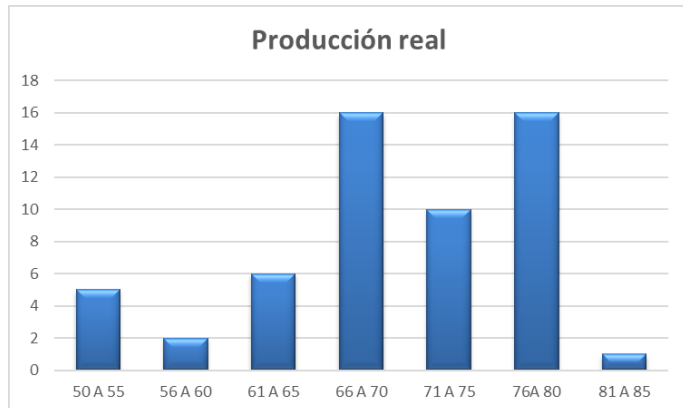
MATRIZ DE DISTRIBUCIÓN DE FRECUENCIA					
Variable Producción real	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa ($f_i = n_i / N$)	Frecuencia relativa acumulada ($F_i = N_i / N$)	Frecuencia porcentual
50 A 55	5	4	0,089	0,071	9%
56 A 60	2	14	0,036	0,250	4%
61 A 65	6	17	0,107	0,304	11%
66 A 70	16	22	0,286	0,393	29%
71 A 75	10	27	0,179	0,482	18%
76A 80	16	30	0,286	0,536	29%
81 A 85	1	56	0,018	1,000	2%
Total	56		1		100%

Nota. SENA (2021)

Con la ayuda del gráfico estadísticos visualizamos la información así:

Figura 2

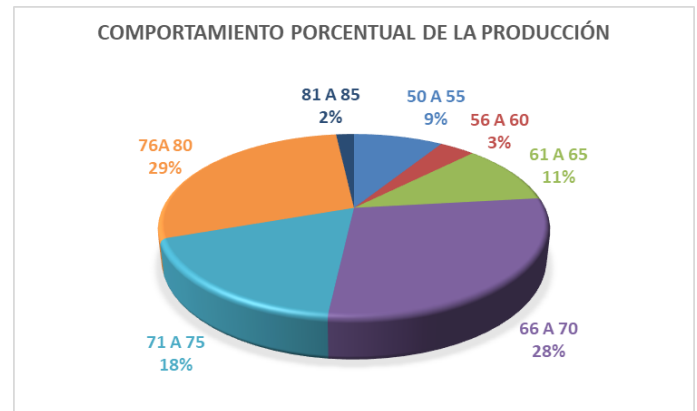
Producción real



Nota. SENA (2021)

Figura 3

Comportamiento porcentual de producción



Nota. SENA (2021)

El comportamiento real de las metas de producción en la semana muestra que, con base en la meta programada (75 unidades por día) en un 29% la planta de producción se movió en un rango promedio entre 66 a 70 unidades, quedando por debajo de las metas programadas por el departamento de planeación. Pero en igualdad de circunstancias un 29% de las metas de producción obtenidas superó la carga programada alojándose en un rango promedio entre 76 a 80 unidades.

En conclusión, la empresa debe realizar ajustes en el aumento de la productividad, para no desestabilizar los indicadores financieros.

Medidas de tendencia central

Las medidas de tendencia central nos orientan hacia la base en que, en cualquier distribución de frecuencias, los datos o valores recopilados luego de tabularlos tienden a una línea central, a concentrarse alrededor de un valor central de la distribución. Orientamos la toma de decisiones a valores mucho más precisos y criterios unificados.

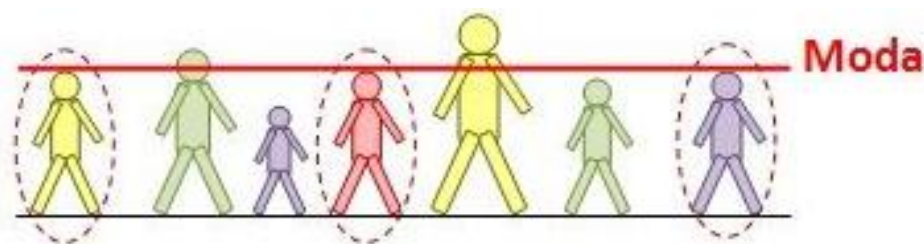
Para entender la naturaleza de la tabulación de los datos, debemos acercarnos lo más rápido posible que tratar de describir el centro de la distribución de las mediciones y la forma como éstas fluctúan alrededor de ese centro.

- **La moda**

Se define como el dato que más se repite, si son dos datos es bimodal, si son 3 es trimodal. La moda de una distribución se define como el valor que presenta la mayor frecuencia. Es comúnmente utilizada como una medida estandarizada, que muestra la tendencia de una opinión.

Figura 4

Moda



Nota. Tomado de Universo Fórmulas (2014)

En algunos eventos o cuadros de distribución sólo hay una moda, pero en otras pueden existir dos o más modas.

Si tomamos 1, 7, 6, 5, 7, 5, 6, 9, 1, 4, 2, 5, 7, 8, 9, 4.

Ordenando tenemos: 1, 1, 2, 4, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 9, 9.

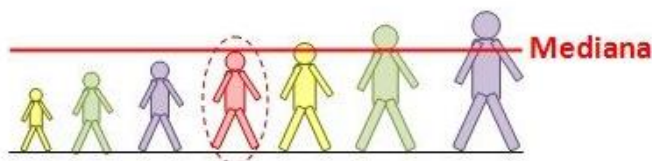
Vemos que tanto el cinco como el siete aparecen con más frecuencia y en tres ocasiones. Es decir, hay dos modas y la distribución es bimodal. Cuando se trabaja con datos agrupados, la moda se refiere como el valor medio del intervalo que constituye la mayor frecuencia.

- **La mediana**

La mediana se define como la medida de tendencia central que divide a cualquier distribución en dos partes iguales. Corresponde a la suma de datos dividido entre la cantidad de estos (promedio). El símbolo utilizado es *Me*.

Figura 5

Mediana



Nota. Tomado de Universo Fórmulas (2014)

En la siguiente distribución: 1, 1, 2, 4, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 9, 9.

La mediana es 5,5 porque se sitúa en el punto que divide la distribución en dos partes iguales, se toman los datos intermedios $\frac{(5+6)}{2}$. Hay el mismo número de casos antes y después del 5.

Cuando hay una distribución con un número par de casos, la mediana es el promedio de los dos valores medios.

La mediana se usa en variables medidas en escala ordinal, intervalo o de razón. Su mayor uso es cuando se tienen muchas observaciones y generalmente se utiliza en distribuciones de ingresos, edades, pesos.

Así, observe la siguiente distribución de eficiencias de producción:

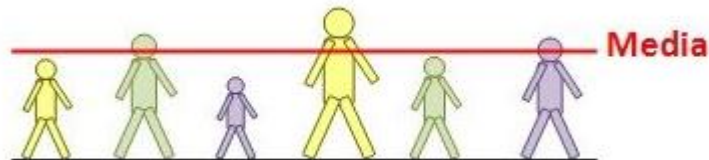
70, 72, 74, 75, 75, 76, 76, 78, 77, 79, 80.

El valor de la mediana es 76, porque, primero al ordenar la distribución de menor a mayor, siendo 11 el total de eficiencias de producción, el dato que aparece en la posición quinta y está en la mitad de la distribución, entonces la mediana será: 76

Cuando se tiene información agrupada, la mediana se define como el valor dentro del intervalo que divide la distribución en dos partes iguales.

- **La media aritmética**

Es la medida de tendencia central más conocida, es fácil de calcular, de gran estabilidad en el muestreo, sus fórmulas permiten tratamiento algebraico (un número algebraico es cualquier número real o complejo). Corresponde a la suma de datos dividido entre la cantidad de estos (promedio). Además, presenta el inconveniente de ser muy sensible a los valores extremos, cuando éstos son demasiado bajos o altos. Se representa así: \bar{X}



Nota. <https://bit.ly/2XtUWMj>

La media aritmética se define como la suma de todos los valores observados dividido por el número de observaciones (n). La fórmula para datos no agrupados es:

Donde $\sum X_i$ corresponde a la sumatoria de todos los valores de la muestra.

La media aritmética de la siguiente serie de datos la obtenemos sumando los datos y dividirlos por (n) datos propuestos: 15,16,17,18,19,20,21,22,23,24,25,26,27,28.

Calculamos la media aritmética así:

$$\sum X_i = \frac{15 + 16 + 17 + 18 + 19 + 20 + 21 + 22 + 23 + 24 + 25 + 26 + 27 + 28}{14}$$

$$\sum X_i = \frac{301}{14} = 21.5$$

El resultado es 21,5 como dato promedio de la serie de datos.

Realmente el porcentaje es un símbolo matemático el cual expresa una cantidad en fracción de 100 partes iguales, trabaja los datos como información, y es la forma más simple para la toma de decisiones, los resultados finales se expresan con el símbolo tanto por ciento (%) y estadísticamente aparece el término indicador.



Veamos con un ejemplo:

Ejemplo porcentajes

MUESTRA	PORCENTUAL %
70	9,40%
71	9,53%
72	9,66%
73	9,80%
74	9,93%
75	10,07%
76	10,20%
77	10,34%
78	10,47%
79	10,60%
745	100%

Nota. SENA (2021)

745 es el 100% y suma total de la muestra y queremos calcular el porcentaje que corresponde a cada uno de ellos, para 70 es el 9,40% de 745, y así sucesivamente.

- **Desviación estándar**

La desviación típica estándar es la raíz cuadrada de la varianza, considerada siempre con signo positivo.

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

S = raíz cuadrada de S²

Para el caso del peso de los 100 estudiantes del colegio la desviación estándar es:

$$S = \text{raíz cuadrada de } 8.52 = 2.92$$

La varianza se expresa siempre en unidades diferentes a las originales, es decir, si la variable se refiere a peso en Kg, al calcular la varianza está dado el peso en Kg² (kilogramos al cuadrado). Es una de las razones por la cual se utiliza la desviación estándar, pues se expresa en las mismas unidades de la variable.