PROYECTO DE AULA ESTADISTICA PARA INGENIERIA

Asignatura:

Estadística Para Ingeniería

2024-2

Docente:

Diana Sirley Guzmán Aguilar

Integrantes:

Isabella Montoya Castellanos

Fecha

2 de Noviembre de 2024

TABLA DE CONTENIDOS

Introducción

La estadística es esencial en ingeniería, ya que proporciona herramientas para el análisis de datos y la toma de decisiones. Este informe se enfoca en tres aspectos clave:

Primero, se analizará el tiempo de producción en dos líneas de fabricación de láminas metálicas con el fin de identificar, mediante las herramientas estadísticas vistas durante el curso, cuál es la línea más eficiente para la empresa.

Segundo, se estudiarán los parámetros de una máquina expendedora de bebidas, utilizando datos para calcular probabilidades y estudiar su funcionamiento.

Finalmente, se presentará una aplicación web que permite obtener resultados de un experimento de lanzar un dado, facilitando el análisis del espacio muestral.

Desarrollo

Punto 1

Considere un proceso de producción de láminas metálicas para escritorios escolares. La empresa cuenta con dos líneas de producción para producir tales láminas, y el área de producción analiza el tiempo (minutos) que tarda cada línea en producir una unidad, para lo cual se tomaron los datos que están en el archivo adjunto en la **hoja llamada BD1.**

a. Calcule los estadísticos muestrales que considere pertinentes para analizar el tiempo que tarda cada línea en producir las láminas. ¿Identifica algún tiempo atípico? Analice e interprete.

	Tiempo Linea 1	Tiempo Linea 2
Media	24.877143	25.082857
Mediana	24.500000	25.040000
Moda	24.610000	24.730000
Desviación Estándar	1.974118	0.625970
Varianza	3.897140	0.391838
Curtosis	0.914326	-0.073031
Coeficiente de Asimetría	1.144372	0.140472
Rango	8.970000	3.040000
Mínimo	21.590000	23.580000
Máximo	30.560000	26.620000
Suma	2437.960000	2633.700000
Cuenta	98.000000	105.000000

Figura 1. Analisis de Estadisticas

Punto a: Análisis Estadístico.

Según el análisis estadístico, la desviación estándar en la Línea 1 es de 1.97, mientras que en la Línea 2 es mucho menor, con un valor aproximado de 0.63. Esto indica que los tiempos en la Línea 1 presentan mayor dispersión, mientras que en la Línea 2 los tiempos son más estables. La varianza también refleja dispersión, con 3.9 en la Línea 1 y solo 0.39 en la Línea 2.

Para poder tener un mejor análisis, realizamos las gráficas de dispersión para cada uno de los Tiempos de Linea:

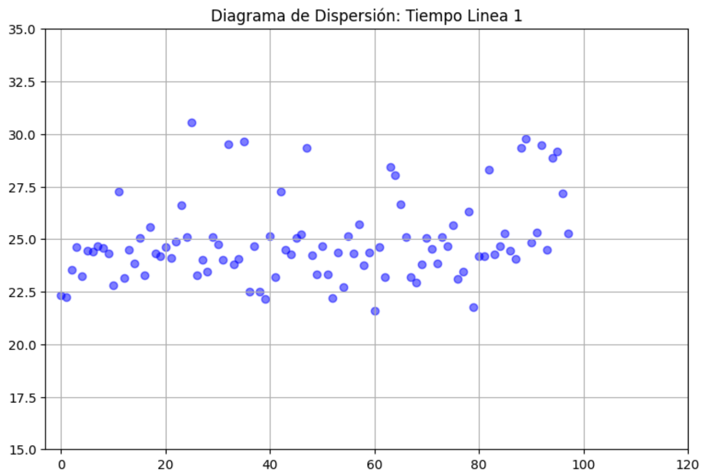


Figura 2. Diagrama de Dispersión Tiempo Linea 1

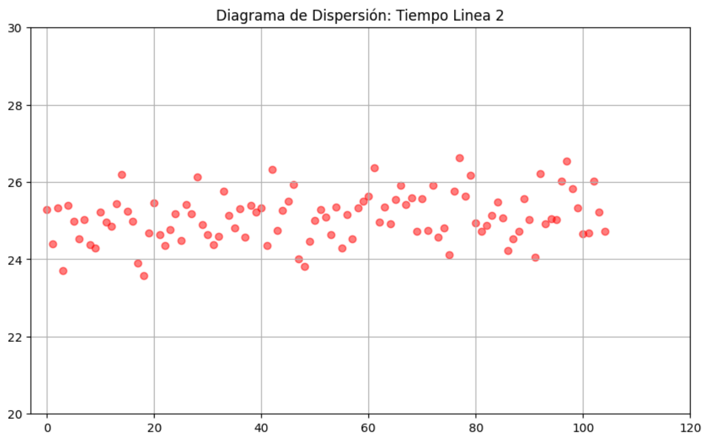


Figura 3. Diagrama de Dispersión Tiempo Linea 2

La Figura 2 y *Figura 3* en este caso no presentan un ajuste en los ejes como podemos observar, así que realizamos un ajuste en los ejes para visualizar mejor la dispersión de los datos en cada Tiempo de Linea y obtuvimos lo siguiente:

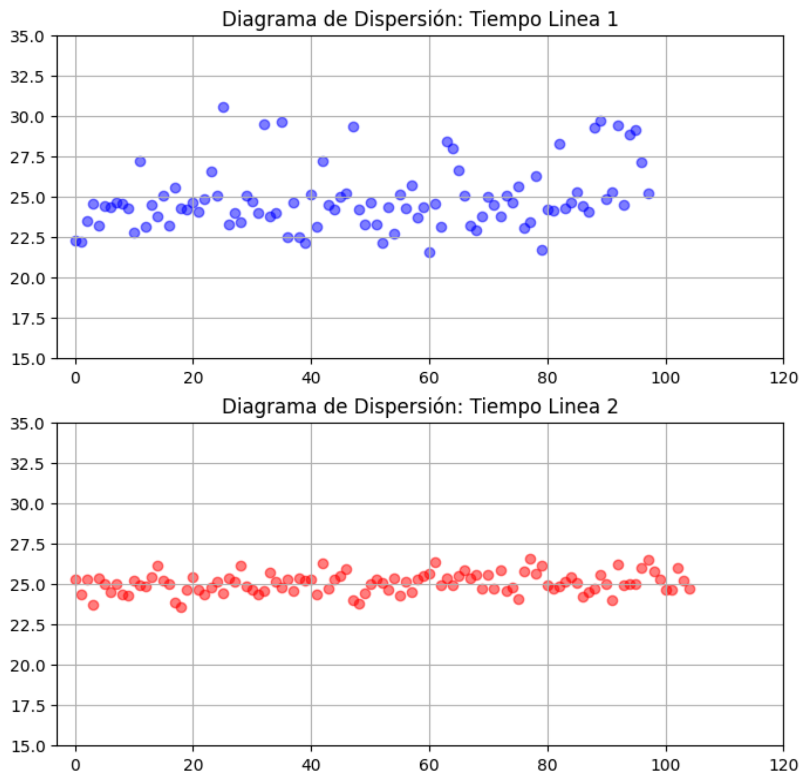


Figura 4. Diagrama de Dispersión Ajustado en Ejes.

Con la Figura 4 podemos confirmar que el Tiempo Línea 1 es el que presenta mayor dispersión que el Tiempo de Linea 2.

La Línea 1 presenta valores atípicos en el diagrama de dispersión de la *Figura 4* debido a puntos de producción muy alejados de la media.

En conclusión, la Línea 1 muestra mayor variabilidad en los tiempos de producción, con valores que incrementan la dispersión, por lo que es probable que existan problemas si llegamos a escoger esta línea.

b. De acuerdo con los datos suministrados, ¿considera que es necesario tomar más tiempos? Analice e interprete.

Punto b:

Con base a lo anteriormente analizado en el *punto a* y las comparaciones resultantes en la *Figura 4*, es recomendable tomar más tiempos para la Línea 1, dado que esta es la que presenta un mayor rango y si tomamos más tiempos para esta línea, podemos obtener un análisis más amplio de cómo están dispersos los datos y si los valores atípicos influyen en estos. Para la Línea 2 no es necesario obtener más tiempos dado que su rango es el más bajo y se han presentado resultados estables.

c. De acuerdo con los datos suministrados, calcule e interprete:

Punto c:

Para resolver el ejercicio hay que tener en cuenta que datos tenemos y qué datos se van a utilizar, para eso recordemos:

Tabla 1

Medidas poblacionales y muestrales

MEDIDA	MUESTRAL	POBLACIONAL
Media	$\overline{X}$	μ
Varianza	S^2	σ^2
Desviación	S	σ

Los datos que obtuvimos del análisis estadístico visualizando la *Figura 1* son muestrales, dado que se realizó un análisis en dos muestras que son el Tiempo de Linea 1 y el Tiempo de Linea 2.

Un intervalo de confianza del 99% para el tiempo promedio de producción por unidad de cada una de las líneas. Compare y analice. Explique para qué le podría servir la construcción de estos intervalos.

Para el promedio de la producción:

Como la varianza poblacional es desconocida y estamos trabajando con un $n \geq 30$ en ambas líneas, usamos la siguiente fórmula:

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1) \qquad \textit{Manera compacta: } \bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} \quad \textbf{(1)}$$

Para la Línea 1:

n = 98 $\quad \overline{X} = 24.877143 \quad S = 1.974118$

Sabemos que la confianza es: $1 - \alpha$

$1 - \alpha = 0.99 \qquad \frac{\alpha}{2} = 0.005 \qquad Z_{0.005} = 2.57$

Reemplazando en la ecuación (1) , por izquierda y por derecha:

$$24.877143 + 2.57\,\frac{1.974118}{\sqrt{98}} = 25.390 \qquad 24.877143 - 2.57\,\frac{1.974118}{\sqrt{98}} = 24.363$$

```
2.5758293035489004
Intervalo de confianza para la media de Línea 1: (24.3634813409332, 25.390804659066802)
```

Figura 5. Resultados en Python I.C para la media Línea 1.

R/ Con una confianza del 99%, se espera que el promedio de producción en la Línea 1 esté entre 24.363 y 25.390.

Para la Línea 2:

$n = 105 \qquad \overline{X} = 25.082857 \qquad S = 0.625970$

$1 - \alpha = 0.99 \qquad \frac{\alpha}{2} = 0.005 \qquad Z_{0.005} = 2.57$

Reemplazando en la ecuación (1) por izquierda y por derecha:

$$25.082857 + 2.57\,\frac{0.625970}{\sqrt{105}} = 25.240 \qquad 25.082857 - 2.57\,\frac{0.625970}{\sqrt{105}} = 24.925$$

```
2.5758293035489004
Intervalo de confianza para la media de Línea 2: (24.925503665728222, 25.24021033427178)
```

Figura 6. Resultados en Python I.C para la media Línea 2.

R/ Con una confianza del 99%, se espera que el promedio de producción en la Línea 2 este entre 24.925 y 25.240.

Análisis:

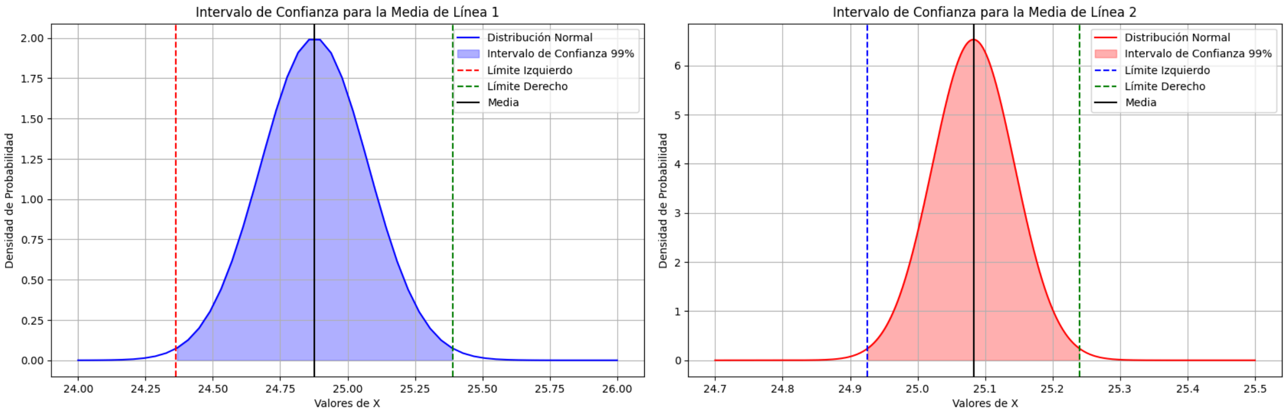


Figura 7. Comparación de I.C Media en línea 1 y línea 2.

Intervalo Promedio Línea 1: 24.363 y 25.390

Intervalo Promedio Línea 2: 24.925 y 25.240

Podemos ver que parte del promedio que se encuentra en la Línea 1 se encuentra también en la Línea 2, como los intervalos son muy cercanos y están uno encima del otro, la diferencia entre ellos es mínima. Pero si comparamos profundamente los intervalos y calculamos el rango de cada uno:

Línea 1 : 25.390 - 24.363 = 1.027

Línea 2 : 25.240 - 24.925 = 0.315

La Línea 1 tiene un intervalo con mayor distancia y un rango más alto, por lo tanto, a mayor distancia de intervalos hay mayor dispersión en los datos. Concluyendo, la Línea 1 presenta más dispersión en su media muestral y es menos eficiente en sus tiempos de producción.

Este análisis se hace con el fin de verificar que tan eficiente es un proceso y qué pasos se deben tomar para mejorarlo.

Un intervalo de confianza del 99% para la desviación del tiempo de producción por unidad de cada una de las líneas de producción. Compare y analice. Explique para qué le podría servir la construcción de estos intervalos.

Para la desviación de la producción:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)} \qquad \frac{(n-1)S^2}{\chi^2_{(\alpha/2, n-1)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2, n-1)}} \qquad \textbf{(2)}$$

Como es la desviación, sacamos raíz cuadrada a la ecuación (2):

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{(\alpha/2, n-1)}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{(1-\alpha/2, n-1)}}} \quad \textbf{(3)}$$

Para la Línea 1:

Los datos que tenemos son los siguientes:

$n = 98 \qquad S^2 = 3.897140$

$1 - \alpha = 0.99 \qquad \frac{\alpha}{2} = 0.005$

Reemplazando en la ecuación (3):

$$\sqrt{\frac{(97)*3.897140}{\chi^2_{(0.005, 97)}}} < \sigma < \sqrt{\frac{(97)*3.897140}{\chi^2_{(1-0.005, 97)}}}$$

Usando la distribución chi-cuadrado obtenemos los siguientes valores para reemplazar, siempre vamos a buscar por cola derecha:

$$\chi^2_{(0.005, 97)} = 136.61 \qquad \chi^2_{(0.995, 97)} = 64.87$$

$$\sqrt{\frac{(97)*3.897140}{136.61}} < \sigma < \sqrt{\frac{(97)*3.897140}{64.87}}$$

Finalmente:

$$1.663 < \sigma < 2.4139$$

```
Valor crítico inferior: 64.87797302850046
Valor crítico superior: 136.61857753108032
Intervalo de confianza para la desviación estándar de Línea 1: (1.663427937865382, 2.413849753054694)
```

Figura 8. Resultados en Python I.C para la Desviación Línea 1.

R/ Con una confianza del 99%, se espera que la desviación de producción en la Línea 1 este entre 1.663 y 2.4139.

Para la Línea 2:

Los datos que tenemos son los siguientes:

$n = 105$ $\qquad S^2 = 0.391838$

$1 - \alpha = 0.99$ $\qquad \frac{\alpha}{2} = 0.005$

Reemplazando en la ecuación (3):

$$\sqrt{\frac{(104)*0.391838}{\chi^2_{(0.005,\,104)}}} < \sigma < \sqrt{\frac{(104)*0.391838}{\chi^2_{(1-0.005,\,104)}}}$$

$$\chi^2_{(0.005,\,104)} = 144.89 \qquad \chi^2_{(0.995,\,104)} = 70.60$$

$$\sqrt{\frac{(104)*0.391838}{144.89}} < \sigma < \sqrt{\frac{(104)*0.391838}{70.60}}$$

Finalmente:

$$0.530 < \sigma < 0.759$$

```
Valor crítico inferior: 70.60635653149876
Valor crítico superior: 144.89132048855896
Intervalo de confianza para la desviación estándar de Línea 2: (0.5303331318937433, 0.7597103524525701)
```

Figura 9. Resultados en Python I.C para la Desviación Línea 2.

R/ Con una confianza del 99%, se espera que la desviación de producción en la Línea 2 este entre 0.530 y 0.759.

Análisis:

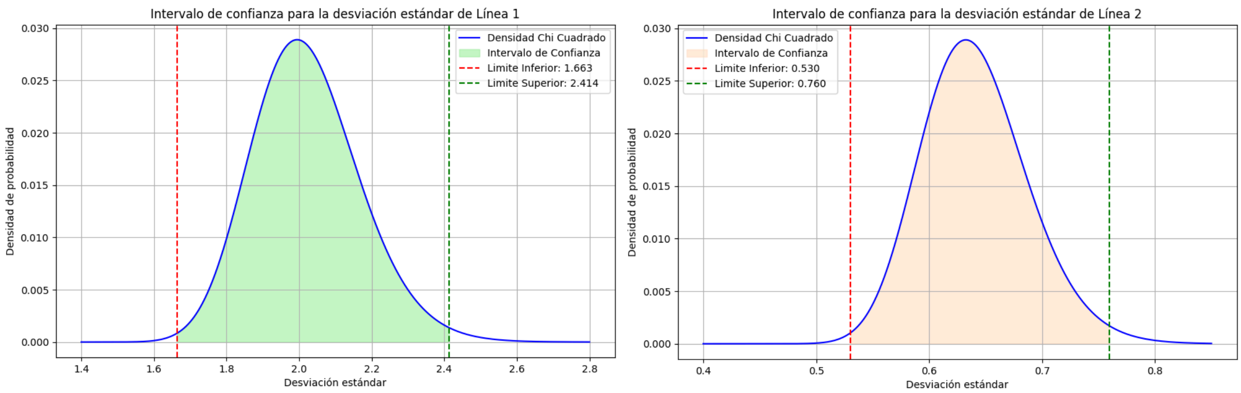


Figura 10. Comparación de I.C Desviación en línea 1 y línea 2.

Intervalo Desviación Línea 1: 1.663 y 2.4139

Intervalo Desviación Línea 2: 0.530 y 0.759

Se puede observar que la Línea 1 presenta mayores valores al de la Línea 2 respecto a la desviación, por lo tanto, la producción de la Línea 1 va a presentar mayor dispersión.

La desviación es importante para calcular la dispersión de los datos. La producción en este caso debe investigar qué causas se presentan para que la Línea 1 se encuentre así de dispersa y con menos confiabilidad. Esto para que defina qué estrategias seguir para mejorar.

Un intervalo de confianza del 99% para la relación de varianzas. Explique para qué le podría servir la construcción de estos intervalos.

Para la relación y comparación de varianzas de la producción:

$$F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}, \qquad \frac{S_1^2}{f_{(\alpha/2, n_1-1, n_2-1)} S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{(\alpha/2, n_2-1, n1-1)} \qquad \textbf{(4)}$$

Los respectivos datos de las muestras son:

Tabla 2

Datos en cada uno de los tiempos de línea.

Tiempo Linea 1	Tiempo Linea 2
$n_1 = 98$	$n_2 = 105$
$S_1^2 = 3.897140$	$S_2^2 = 0.391838$
$S_1 = 1.974118$	$S_2 = 0.625970$

Reemplazando en la ecuación (4):

$$1 - \alpha = 0.99 \qquad \frac{\alpha}{2} = 0.005$$

$$\frac{S_1^2}{S_2^2 f_{(\alpha/2, n_1-1, n_2-1)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{(\alpha/2, n_2-1, n_1-1)}$$

$$\frac{3.897140}{0.391838\, f_{(0.005, 98-1, 105-1)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{3.897140}{0.391838} f_{(0.005, 105-1, 98-1)}$$

Encontramos el valor con la distribución F, siempre en cola derecha:

$$f_{(0.005, 97, 104)} = 1.6758 \qquad f_{(0.005, 104, 97)} = 1.6826$$

Reemplazando los valores:

$$\frac{3.897140}{0.391838(1.6758)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{3.897140}{0.391838} 1.6826$$

$$5.93 < \frac{\sigma_1^2}{\sigma_2^2} < 16.73$$

```
Valor critico inferior: 1.6758289748631359
Valor critico superior: 1.682623113573861
Intervalo de confianza para la relación de varianzas: (5.934850199050125, 16.735022741115554)
```

Figura 11. Resultados en Python I.C para la razón de Varianzas.

Análisis:

Tenemos que la relación de las varianzas se encuentran dentro de intervalos donde tienen un valor mayor a 1, entonces:

$$\frac{\sigma_1^2}{\sigma_2^2} > 1 = \sigma_1^2 > \sigma_2^2$$

Realizamos este análisis con la relación de varianzas para saber qué Línea es más confiable y mucho mejor, podemos ver que la varianza de la Línea 1 es mayor que la de la Línea 2, por lo tanto la Línea más confiable sería la 2. También podemos afirmar que las varianzas son diferentes ya que el 1 no se encuentra en el intervalo de confianza.

d. Analice la normalidad del tiempo de producción por unidad en cada una de las líneas, usando la prueba chi cuadrado.

Punto d

Realizaremos una Prueba de Bondad de Ajuste, con las siguientes hipótesis:.

Para la Línea 1:

$$H_0: Tiempos\ de\ la\ Línea\ 1\ provienen\ de\ una\ Población\ Normal$$

$$H_a: Tiempos\ de\ la\ Línea\ 1\ NO\ provienen\ de\ una\ Población\ Normal$$

Así se ven los datos ajustados a una distribución normal, con una confianza del 95%

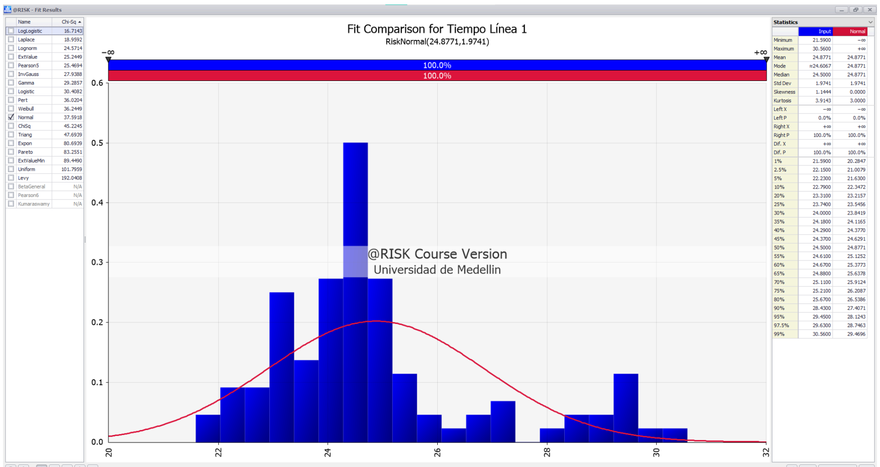
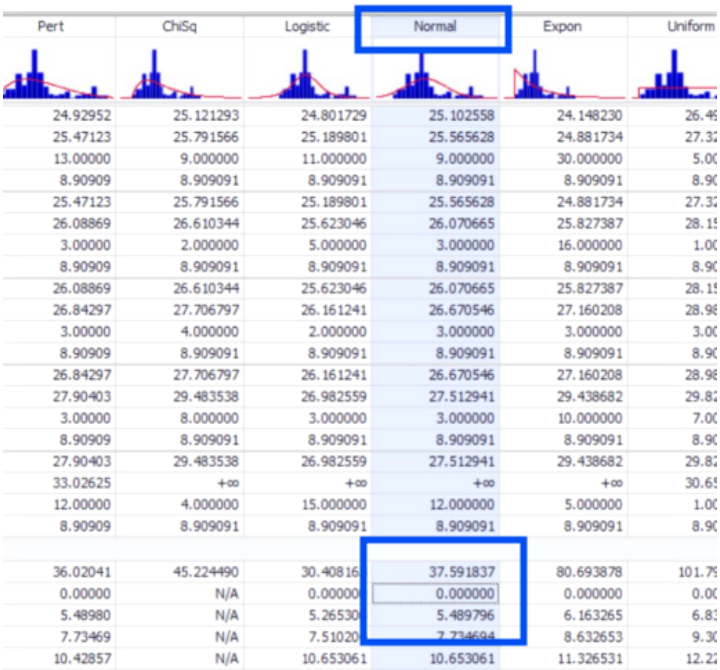


Figura 12. Distribución Normal con confianza del 95%



Pert	ChiSq	Logistic	Normal	Expon	Uniform
24.92952	25.121293	24.801729	25.102558	24.148230	26.4
25.47123	25.791566	25.189801	25.565628	24.881734	27.3
13.00000	9.000000	11.000000	9.000000	30.000000	5.0
8.90909	8.909091	8.909091	8.909091	8.909091	8.9
25.47123	25.791566	25.189801	25.565628	24.881734	27.3
26.08869	26.610344	25.623046	26.070665	25.827387	28.1
3.00000	2.000000	5.000000	3.000000	16.000000	1.0
8.90909	8.909091	8.909091	8.909091	8.909091	8.9
26.08869	26.610344	25.623046	26.070665	25.827387	28.1
26.84297	27.706797	26.161241	26.670546	27.160208	28.9
3.00000	4.000000	2.000000	3.000000	3.000000	3.0
8.90909	8.909091	8.909091	8.909091	8.909091	8.9
26.84297	27.706797	26.161241	26.670546	27.160208	28.9
27.90403	29.483538	26.982559	27.512941	29.438682	29.8
3.00000	8.000000	3.000000	3.000000	10.000000	7.0
8.90909	8.909091	8.909091	8.909091	8.909091	8.9
27.90403	29.483538	26.982559	27.512941	29.438682	29.8
33.02625	+∞	+∞	+∞	+∞	30.6
12.00000	4.000000	15.000000	12.000000	5.000000	1.0
8.90909	8.909091	8.909091	8.909091	8.909091	8.9
36.02041	45.224490	30.40816	37.591837	80.693878	101.7
0.00000	N/A	0.00000	0.000000	0.000000	0.0
5.48980	N/A	5.26530	5.489796	6.163265	6.8
7.73469	N/A	7.51020	7.734694	8.632653	9.3
10.42857	N/A	10.653061	10.653061	11.326531	12.2

Figura 13. Normal con prueba Chi-cuadrado Línea 1

De la tabla en la *Figura 13* se puede sacar el Estadístico de Prueba y el Valor P, usando la prueba Chi Cuadrado, para una normal.

$$Estadístico\ de\ Prueba\ = 37.59, \quad Valor\ P\ = 0 \quad , \alpha\ = 0.05$$

Para rechazar H_0 , entonces $Valor\ P\ <\ \alpha$. En este caso tenemos que $0 < 0.05$, por lo que rechazamos H_0 .

Con una confianza del 95%, los tiempos de la Línea 1 NO provienen de una población normal.

Ahora, la distribución con un mayor Valor P, es la que mejor se ajusta a los datos de la Línea 1, en este caso, la distribución Log Logistic obtuvo un $Valor\ P\ = 0.025$, siendo el mayor de todos los valores P.

Así se ve la gráfica de los Tiempos de la Línea 1, usando una distribución Log Logistic.

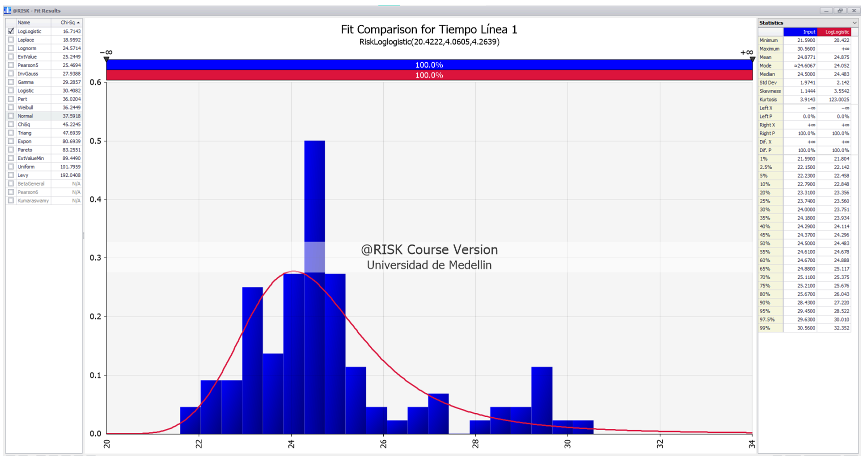


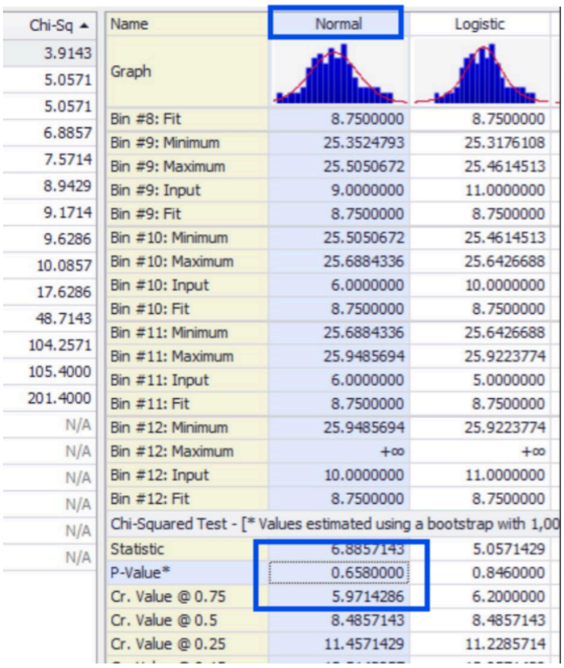
Figura 14. Distribución Log Logistic Línea 1

Para la Línea 2:

$$H_0 \text{: } Tiempos\ de\ la\ Línea\ 2\ provienen\ de\ una\ Población\ Normal$$

$$H_a \text{: } Tiempos\ de\ la\ Línea\ 2\ NO\ provienen\ de\ una\ Población\ Normal$$

Con una confianza del 95%, estos son los datos para la Línea 2.



Chi-Sq ▲
3.9143
5.0571
5.0571
6.8857
7.5714
8.9429
9.1714
9.6286
10.0857
17.6286
48.7143
104.2571
105.4000
201.4000
N/A
N/A
N/A
N/A
N/A
N/A

Name	Normal	Logistic
Graph		
Bin #8: Fit	8.7500000	8.7500000
Bin #9: Minimum	25.3524793	25.3176108
Bin #9: Maximum	25.5050672	25.4614513
Bin #9: Input	9.0000000	11.0000000
Bin #9: Fit	8.7500000	8.7500000
Bin #10: Minimum	25.5050672	25.4614513
Bin #10: Maximum	25.6884336	25.6426688
Bin #10: Input	6.0000000	10.0000000
Bin #10: Fit	8.7500000	8.7500000
Bin #11: Minimum	25.6884336	25.6426688
Bin #11: Maximum	25.9485694	25.9223774
Bin #11: Input	6.0000000	5.0000000
Bin #11: Fit	8.7500000	8.7500000
Bin #12: Minimum	25.9485694	25.9223774
Bin #12: Maximum	+∞	+∞
Bin #12: Input	10.0000000	11.0000000
Bin #12: Fit	8.7500000	8.7500000
Chi-Squared Test - [* Values estimated using a bootstrap with 1,00		
Statistic	6.8857143	5.0571429
P-Value*	0.6580000	0.8460000
Cr. Value @ 0.75	5.9714286	6.2000000
Cr. Value @ 0.5	8.4857143	8.4857143
Cr. Value @ 0.25	11.4571429	11.2285714

Figura 15. Normal con prueba Chi-cuadrado Línea 2

De esta tabla en la *Figura 15* se puede sacar el Estadístico de Prueba y el Valor P, usando la prueba Chi Cuadrado, para una normal.

$$Estadístico\ de\ \ Prueba\ = 6.88, \qquad Valor\ P\ = 0.658, \alpha\ = 0.05$$

Para NO rechazar H_0 , entonces $Valor\ P\ > \alpha$. En este caso tenemos que 0.658 > 0.05, por lo que NO rechazamos H_0 .

Con una confianza del 95%, los tiempos de la Línea 2 provienen de una población normal.

Ahora, la distribución con un mayor Valor P, es la que mejor se ajusta a los datos de la Línea 2, en este caso, la distribución Logistic obtuvo un $Valor\ P\ = 0.846$, siendo el mayor de todos los valores P.

Así se ve la gráfica de los Tiempos de la Línea 2, usando una distribución Logistic.

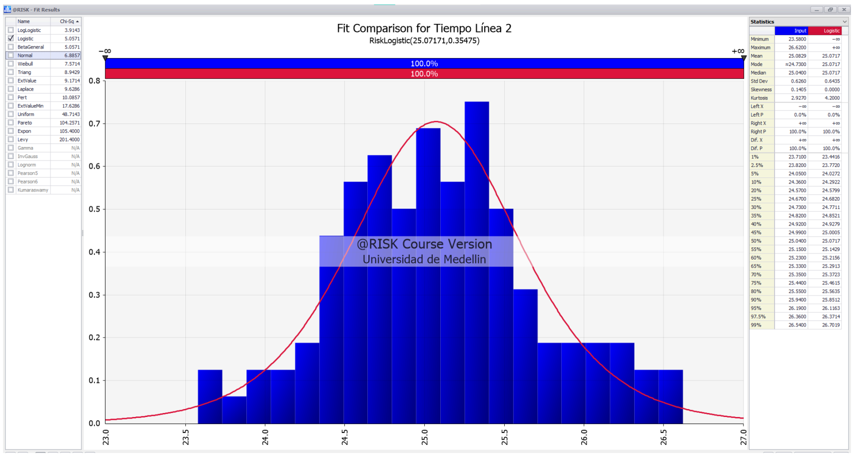


Figura 16. Distribución Logistic Línea 2

e. El proveedor de las líneas de producción indica que una pieza debe ser producida entre 24 ± 0.06 minutos ¿Cuál de las líneas se ajusta mejor al tiempo que indica el proveedor? Explique analíticamente su respuesta

Para determinar cuál de las dos líneas se ajusta mejor al tiempo especificado por el proveedor haremos un análisis comparativo entre ambas.

Intervalo indicado por el proveedor:

$$24\ \pm\ 0.06\ \Rightarrow\ [23.94,\ 24.06]$$

Tabla 3

Análisis comparativo

	Línea 1	Línea 2
Media muestral	24.877	25.082
Intervalo de confianza para la media poblacional	[24.363, 25.39]	[24.925, 25.24]
Diferencia con el tiempo especificado	$\lvert 24.877143-24 \rvert=0.877$	$\lvert 25.082-24 \rvert=1.082$

Según el análisis mostrado en la *Tabla 3* podemos observar que aunque los intervalos de tiempo promedio de las dos líneas están fuera del intervalo indicado por el proveedor, la que mejor se ajusta es la línea 1, ya que su tiempo promedio es más cercano al tiempo promedio especificado.

f. ¿Cuál de las dos líneas es más productiva?

Punto f

Al analizar estadísticamente el tiempo en cada línea tarda en producir una lámina, podemos observar que aunque la línea 1 muestra en promedio menos tiempo empleado en producir una unidad que la línea 2, bajo este criterio, podemos decir que la línea 1 sería la más productiva, sin embargo, también muestra una varianza notablemente mayor que la línea 2, lo cual hace que la línea 2 aunque muestre registros ligeramente más altos es la más confiable, además, en la línea 1 se observa tendencia a valores atípicos altos lo que puede afectar a la producción de la fabricación a largo plazo.

Punto 2

Una máquina que expende bebidas gaseosas está calibrada de modo que descargue un promedio de μ mililitros por vaso. Si la cantidad de líquido está distribuida normalmente con una desviación estándar igual a σ ml. Tanto μ como σ son desconocidos, pero se cuenta con una muestra de doscientas observaciones para estimarlos. Utilice los estadísticos X y S como estimadores de μ y σ, respectivamente, para responder los siguientes literales. Recuerde que:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2$$

a. ¿Qué porcentaje de vasos contendrá menos de 153 ml?

Punto a:

$X = Cantidad\ de\ líquido\ en\ ml \qquad\qquad n = 200$

Sabemos que $X \sim N(\mu, \sigma^2)$

Entonces podemos encontrar quien es $\overline{X}$ y S^2:

$$\overline{X} = \frac{1}{200}\sum_{i=1}^{200} x_i \qquad\qquad S^2 = \frac{1}{199}\sum_{i=1}^{200} (x_i - \overline{x})^2$$

Con los datos dados en el archivo, logramos obtener su respectiva media muestral y varianza muestral.

$\overline{X} = 149.97 \qquad\qquad S^2 = 1.3712$

Por lo tanto:

$X \sim N(149.97, 1.3712) \qquad\qquad S = 1.1709$

$P(X < 153) = 0.995$

R/ El porcentaje de vasos que contendrá menos de 153 ml es de un 99.5%.

b. ¿Cuál es la probabilidad de que un vaso contenga entre 153 y 157 ml?

Punto b

Usando la siguiente propiedad:

$F(b) - F(a) = P(X \leq a) - P(X \leq b)$

$P(153 < X < 157) = P(X \leq 157) - P(X \leq 153)$

$= 1 - 0.99591 = 0.00489$

R/ La probabilidad de que un vaso contenga entre 153 y 157 ml es de 0.00489 es decir, un aproximadamente un 0.49%.

c. Si se usan vasos de 160 ml, ¿cuántos de los siguientes 850 vasos se derramarán?

Punto c

La probabilidad de que un solo vaso se derrame es de:

$$P(X > 160) = 0$$

Para obtener la cantidad de vasos que se van a derramar, multiplicamos por la cantidad de vasos.

$$0 \times 850 = 0$$

R/ Si se usan vasos de 160 ml, de los 850 vasos dados ninguno se va a derramar.

d. ¿Bajo qué valor estará el 5% de los vasos con menos contenido?

Punto d

$$P(X < x) = 0.05 \qquad x = 148.04$$

R/ Un 5% de los vasos con menos contenido se van a encontrar bajo un valor de 148.04.

Punto 3

Investigue para qué sirve la estadística específicamente en la ingeniería que estudia y realice una aplicación con datos reales de alguna de las aplicaciones que usted investigó.

Estadística en Ingeniería de Sistemas

La importancia de la estadística en esta rama de la ingeniería es que facilita el análisis y la visualización de grandes cantidades de datos, con la estadística se hace de manera más rápida, dado que se extrae mucha información para la realización de algoritmos de Machine Learning y la toma de decisiones al momento en que se realizan análisis de datos en las empresas.

Software

Con nuestro software buscamos facilitar la visualización de espacios muestrales para lanzamiento aleatorio de dados, logrando una mayor comprensión en el comportamiento de todos los posibles valores que se pueden obtener en el lanzamiento de N dados. Esta herramienta se puede utilizar para mejorar la pedagogía en la enseñanza de la estadística, ayudando a los profesores a tener una manera de mostrar más fácilmente cómo se ve un espacio muestral, y ayudando a los estudiantes para que la puedan utilizar para resolver problemas propuestos en clase. Para lograr mostrar todos los posibles valores, utilizamos la regla de la multiplicación para primero conocer cuántos posibles resultados existen, y entonces con un ciclo ir sumando en cada dado hasta conseguir todos los posibles resultados.

El análisis y la presentación se encuentran en los siguientes enlaces:

PROYECTO DE AULA ESTADISTICA.ipynb PRESENTACIÓN CANVA

Software Desplegado:

https://multiplicationrule.netlify.app

Conclusiones

Por medio de la primera parte de este estudio estadístico hemos podido demostrar que la línea 2 de producción es la línea que presenta registros más constantes y confiables, aunque la línea 1 presenta tiempos ligeramente más rápidos en promedio, la línea 2 sigue siendo una buena opción debido a que presenta muchos menos valores atípicos y una menor dispersión.

En la segunda parte se ha mostrado que la probabilidad de que un vaso sea llenado con menos de 153 ml es muy alta (99.5%) mientras que la probabilidad de que sea llenado con más de 160 ml es, para el caso de nuestra muestra prácticamente cero, por lo que podemos estar confiados de que un vaso diseñado para contener menos de 160 ml casi nunca se derramará.

El software diseñado para la visualización de espacios muestrales puede llegar a ser una herramienta útil para estudiantes y profesores. Con esta herramienta se puede ilustrar conceptos estadísticos, observando cuán grande puede llegar a ser un espacio muestral para una situación tan simple como un lanzamiento de dados.

Bibliografía

Investigación Científica. (2023, 4 de enero). *Importancia de la estadística aplicada en la informática*. https://guidocutipa.blog.bo/investigacion/importancia-estadistica-informatica/

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2014). *Estadística matemática con aplicaciones* (7.ª ed.). Cengage Learning.

Devore, J. L. (2011). *Probabilidad y estadística para ingeniería y ciencias* (7.ª ed.). Cengage Learning.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & van der Walt, S. J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261-272.https://scipy.org/

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.https://matplotlib.org/