Ajuste de Distribuciones

Práctica 4

Rafael Rus Rus Grupo 82 100363907

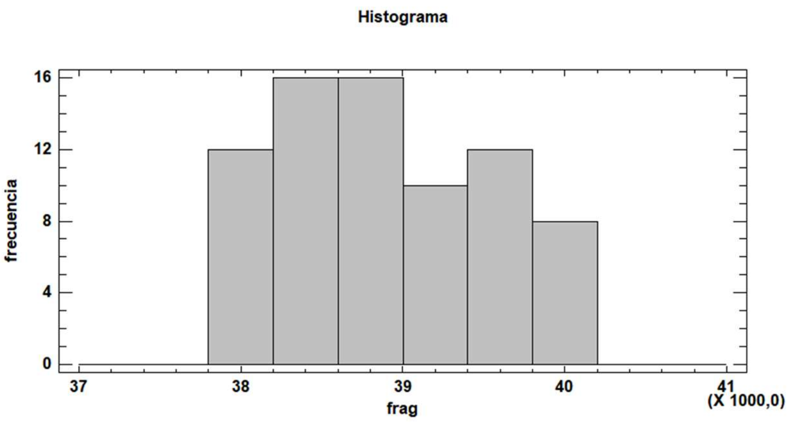
Ruichao Lu Grupo 82 100363792

Ejercicio 2

a) Antes de contestar a la pregunta pedida en el apartado a, debemos realizar el análisis descriptivo de los datos (medidas características e histograma) de la variable "frag":

Resumen Estadístico para frag

Recuento	74
Promedio	38874,4
Desviación Estándar	649,33
Coeficiente de Variación	1,67033%
Mínimo	37844,0
Máximo	39984,0
Rango	2140,0
Sesgo	0,0999263
Curtosis	-1,14922



Como podemos observar en el histograma (en el que se han utilizado 10 clases), los datos de la muestra son unimodales y simétricos (sesgo o coeficiente de Fisher ≈ 0). La zona de la moda (localizada justo en el 37900) tiene un apuntamiento en forma de campana. Por tanto, debido a la simetría, podemos afirmar que la variable "frag" puede distribuirse como una distribución normal; sin embargo, tenemos la posibilidad de intentar ajustar una normal a una transformación de los datos para que sea mejor.

b) Para que el ajuste a la normal sea mejor (tal y como hemos dicho en el apartado anterior), debemos realizar una transformación del tipo, por ejemplo, raíz cuadrada:

$$y = \sqrt{x}$$

en la que se comprime la escala en los valores altos y se expande en los valores bajos.

Sesgo	0,0892974

Al realizar la transformación podemos observar que el sesgo ≈ 0, por lo que la distribución sigue siendo simétrica, y el ajuste a la normal es mejor.

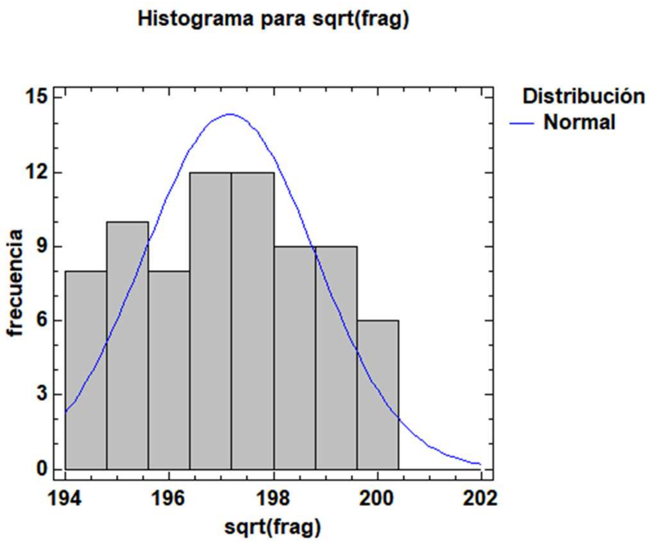
c) Estimación de los parámetros del modelo escogido:

Datos/Variable: sqrt(frag)

74 valores con rango desde 194,535 a 199,96

Distribuciones Ajustadas

Normal
media = 197,159
desviación estándar = 1,64611



El modelo escogido será una Normal de media 197.159 y desviación típica 1.64611

Pruebas de Bondad-de-Ajuste para sqrt(frag)
Prueba Chi-Cuadrada

	Límite Inferior	*Límite Superior*	*Frecuencia Observada*	*Frecuencia Esperada*	*Chi-Cuadrada*
menor o igual		194,0	0	2,03	2,03
	194,0	194,8	8	3,58	5,44
	194,8	195,6	10	7,10	1,19
	195,6	196,4	8	11,14	0,89
	196,4	197,2	12	13,88	0,25
	197,2	198,0	12	13,72	0,21
	198,0	198,8	9	10,75	0,29
	198,8	199,6	9	6,69	0,80
mayor	199,6		6	5,11	0,16

Chi-Cuadrada = 11,2622 con 6 g.l. Valor-P = 0,080603

Prueba de Kolmogorov-Smirnov

	Normal
DMAS	0,0759515
DMENOS	0,0841476
DN	0,0841476
Valor-P	0,671228

Anderson-Darling A^2

	Normal
A^2	0,788759
Forma Modificada	0,788759
Valor-P	>=0.10

Para hacer la estimación de los parámetros del modelo se debe realizar el test de la chi-cuadrado, cuyo resultado se resume en:

- Chi-Cuadrada = 11.622 -> estadístico que resume la discrepancia entre el histograma y la curva de la normal. Dado que el número de clases proporciona unas frecuencias esperadas y observadas superiores a 5 (observado en la mayoría de los intervalos), el ajuste de nuestros datos al modelo es correcto.

- g.l = 6 -> grados de libertad de la distribución chi-cuadrado (6 = 9 [clases] – 2 [parámetros] – 1).

- Valor-P = 0.080603 -> área que queda a la derecha del estadístico calculado en la distribución (11.622).

Dado que el p-valor es mayor que 0.05, consideramos que el ajuste es suficientemente bueno, y que el modelo elegido puede usarse como modelo para la población.

Concluimos, pues, que la normal X ~ N (197.159, 1.64611) es un modelo muy razonable para explicar la distribución de nuestros datos.

d) Utilizando el modelo estimado en el apartado anterior (en el cual vimos que la raíz cuadrada del tiempo se ajusta bien a la Normal) podemos calcular la probabilidad de que el tiempo sea mayor de 39000 unidades, el cual es equivalente a $\sqrt{39000} = 197.484$ unidades.

Áreas de Cola para sqrt(frag)
Distribución Normal

X	*Área Cola Inferior (<)*	*Área Cola Superior (>)*
197,484	0,578255	0,421745

$$P(X > 197.484) = 0.421745 = 42.17\ \%$$