Problema 4 Guía 0.

La siguiente tabla da la distribución del tiempo de duración en segundos de 1000 llamadas telefónicas :

Duración (marca de clase)	30	60	90	120	150	180	210	240	270	300
Frecuencia	6	28	88	180	247	260	133	42	11	5

1. Representar el histograma y polígono de frecuencias.
2. Obtener la tabla representar el polígono de frecuencias acumuladas.
3. Calcular la media , mediana, modo y desvío standard muestral s.
4. los porcentajes de llamadas cuya duración pertenece a los intervalos centrados en la media y de semiamplitud 1) s 2) 2 s 3) 3 s.
5. Calcular el porcentaje de llamadas cuya duración supera los 3 minutos (Rta: 32 %)

Nota: La marca de clase es el punto medio del intervalo de clase (que en este ejemplo tiene una amplitud de 30 seg). Para responder c) y d) es conveniente realizar un gráfico “linealizado” de las frecuencias acumuladas (en el que la frecuencia acumulada en cada intervalo crece linealmente).

Respuesta:

La información disponible corresponde a la tabla de frecuencias construida a partir de la duración en segundos de 1000 llamadas telefónicas. En este caso el rango de la duración de las llamadas fue subdividido en 10 intervalos de igual longitud (30 segundos). El primer intervalo corresponde a llamadas cuya duración es de 15 a 45 (y son 6 llamadas), el segundo intervalo es (45, 75] y se registraron 28 llamadas y así sucesivamente.

El siguiente código en *Octave/Matlab* genera el histograma (Fig. 1) y el polígono de frecuencias (Fig. 2):

```
d=[30 60 90 120 150 180 210 240 270 300]; % marcas de clase
f=[6 28 88 180 247 260 133 42 11  5]; % frecuencias absolutas
bar(d,f,1) % genera el histograma con barras contiguas
% el vector dd tiene los extremos de los intervalos de clase
dd =[ 0    30    60    90   120   150   180   210   240   270   300   330];
% el vector ff tiene las ordenadas para el poligono de frecuencias
ff=[0 6 28 88 180 247 260 133 42 11  5 0];
e =[ 15  15  45  45  75  75  105 105  135  135 165  165];
e=[e 195 195  225 225 255 255 285 285 315 315];
fff =[ 0  6  6  28  28 88  88 180 180 247 247 260 260 133 133 42 42  11 11 5 5  0];
% los vectores e y fff tienen las coordenadas de la poligonal
% que limita el histograma
% en el grafico que sigue se superponen el histograma y
% el poligono de frecuencias
plot(dd,ff,'-o',e,fff,'-')
```

La tabla con la información necesaria para construir el polígono de frecuencias acumuladas es la siguiente:

Duración (marca de clase)	15	45	75	105	135	165	195	225	255	285	315
Frecuencia	0	6	34	122	302	549	809	942	984	995	1000

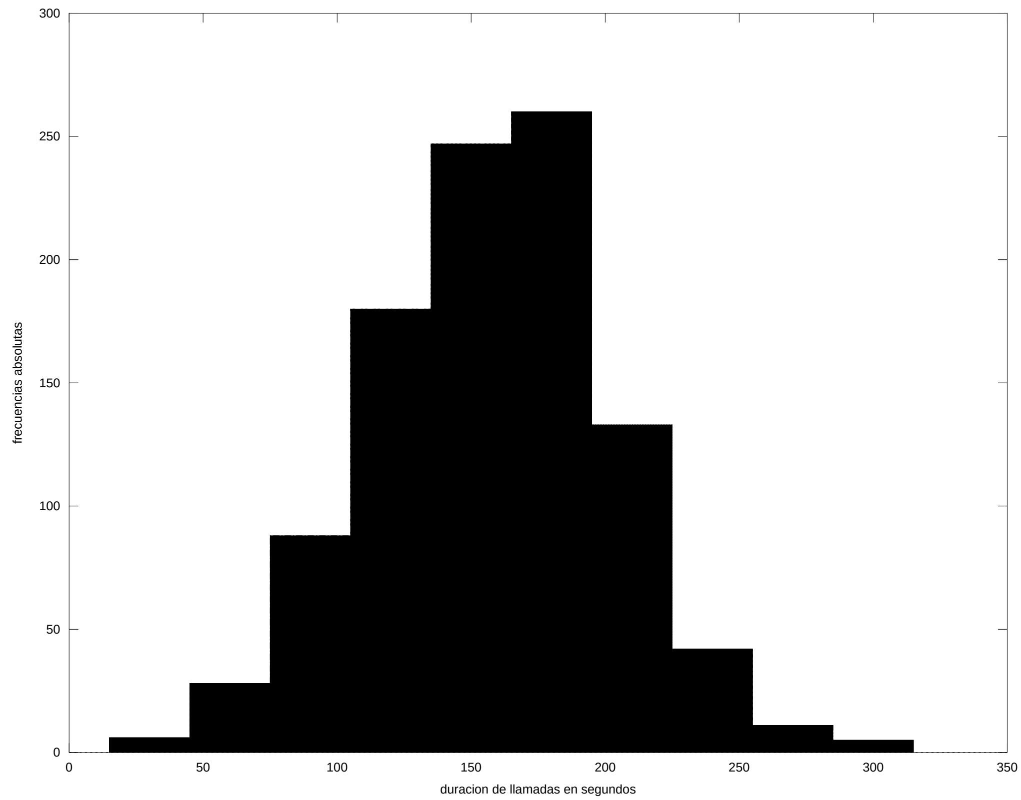

Figura 1: Histograma de frecuencias

El siguiente código en *Octave/Matlab* genera el histograma de frecuencias acumuladas y el polígono de estas frecuencias:

```
x=[15 15 45 45 75 75 105 105  135 135 165 165  195 195 225 225 255 255 285 285 315];
F=[0 6 6 34 34 122 122 302 302 549 549 809 809 942 942 984 984 995 995 1000 1000];
xx=[15 45 75 105 135 165 195 225 255 285 315 ];
FF=[0 6 34 122 302 549 809 942 984 995 1000];
plot(x,F,'-',xx,FF,'-o')
ylabel('frecuencias acumuladas absolutas')
xlabel('duracion de llamadas en segundos')
title('Histograma y poligono de frecuencias acumuladas')
```

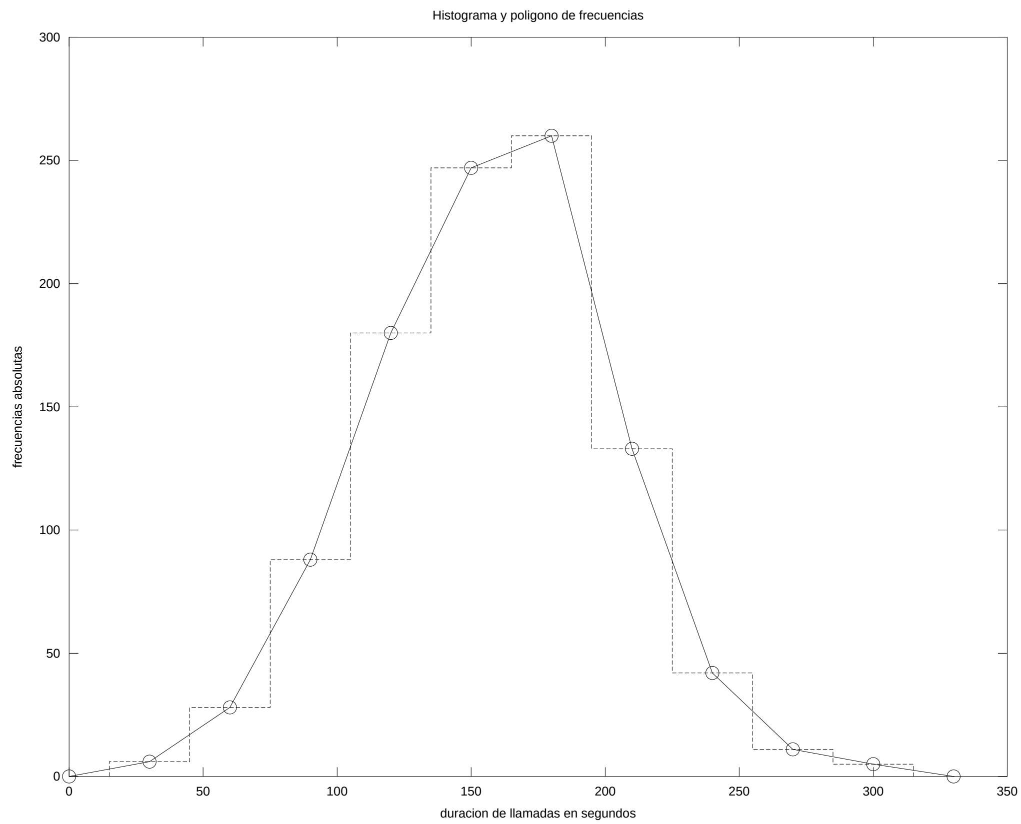

Figura 2: Histograma y polígono de frecuencias

Media y varianza muestrales con datos agrupados

La media de la muestra analizada, usando datos agrupados, se calcula de la siguiente manera:

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{m} m_k \, f_k,$$

donde m_k y f_k son, respectivamente, la marca de clase y la frecuencia absoluta del intervalo de clase k. El tamaño de la muestra es n, con $n = \sum_{k=1}^{m} f_k$.
La varianza muestral viene dada por

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{m} (m_k - \overline{x})^2 \, f_k.$$

Con los datos de esta tabla:

$$\overline{x} = 157.71 \text{ seg},$$

$$s \approx 45.37 \text{ seg}.$$

En este caso la distribución de frecuencias tiene un único modo. Una opción es asignarle al modo el valor del punto medio del intervalo al que corresponde la máxima frecuencia (el sexto intervalo) o sea la abscisa del modo del polígono de frecuencias. Así resulta que $M = 180$ seg, el punto medio

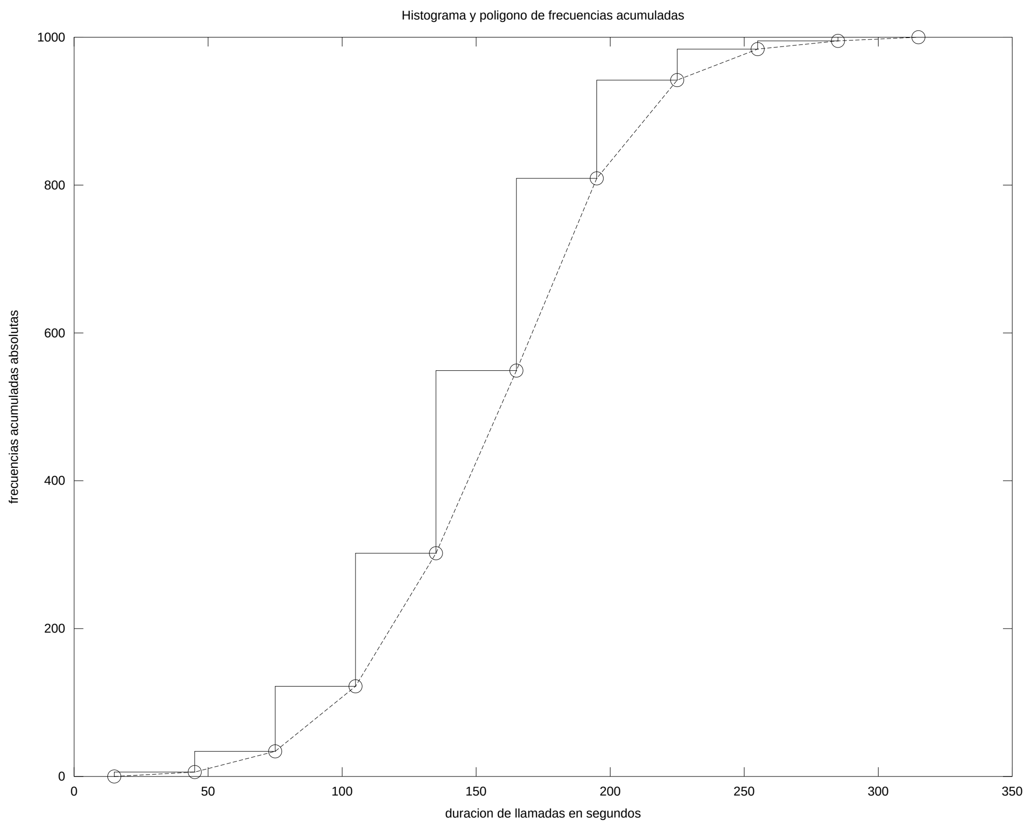

Figura 3: Histograma y polígono de frecuencias acumuladas

de ese intervalo. Otra opción utilizada para determinar el modo M tiene en cuenta las frecuencias de los dos intervalos contiguos (por izquierda f_I y por derecha f_D) y realizar una asignación proporcional de la forma $(M - L_I)\,(f_M - f_D) = (L_D - M)\,(f_M - f_I)$, donde L_I y L_D son los extremos izquierdo y derecho del intervalo al que pertenece M. En este caso (y en segundos) son $L_I = 165$ y $L_D = 195$ en tanto que $f_I = 247$, $f_D = 133$ y $f_M = 260$. Reemplazando, resulta, $M \approx 167.79$ seg.

Para determinar un valor de la mediana se puede realizar una interpolación lineal sobre el polígono de frecuencias acumuladas para obtener la abscisa a la que corresponde la frecuencia acumulada $n/2$ (en este caso 500). Esa frecuencia acumulada pertenece al intervalo (135, 165) y la frecuencia acumulada a cada extremos es 302 y 549, respectivamente. Entonces la mediana m_e resulta $m_e = 135 + 30\,\frac{500-302}{549-302} \approx 159.05$ seg.

Por interpolación lineal en la tabla de frecuencias acumuladas se puede responder a preguntas relacionadas con la proporción de llamadas en determinado intervalo de valores de la duración de ellas. Los intervalos centrados en la media y de semiamplitud s, 2 s y 3 s son los intervalos (112.34 203.08), (66.97 248.45) y (21.60 293.82) respectivamente. Si $P(x)$ es la función interpoladora lineal por tramos de las frecuencias acumuladas entonces la proporción de llamadas cuya duración pertenece al intervalo (a, b) viene dada por $(P(b) - P(a))/n$. El gráfico de P es el polígono de frecuencias acumuladas, la interpolación lineal supone que la densidad de datos es constante que es la suposición acorde a disponer de datos agrupados.

La frecuencia acumulada interpolada para 112.34 se obtiene considerando los pares (105, 122) y (135, 302) y generando $P(112.34) = 122 + (302 - 122)(112.34 - 105)/30 \approx 166.04$. Por otro lado la frecuencia acumulada interpolada para 203.08 se obtiene considerando los pares (195, 849) y (225, 942) y generando $P(203.8) = 195 + (942 - 809)(203.08 - 195)/30 \approx 844.82$. Asi, entonces, la proporción de llamadas cuya duración pertenece al intervalo (112.34 203.08) es $(844.82 - 166.04)/1000 \approx 0.68$ o sea 68 %. En igual forma se pueden obtener los porcentajes de datos en cada uno de los otros dos intervalos.

El porcentaje de llamadas cuya duración supera los 3 minutos corresponde a evaluar $(1 - P(180))/1000$. Es sencillo verificar que $P(180) = (809 + 549)/2 = 679$ de donde el porcentaje de llamadas cuya duración supera los 3 minutos es 32.1 %.