

HISTOGRAMAS DE FRECUENCIAS

SIMULACION DE SISTEMAS

3007331

Adaptado de capítulo 3 Allen (2011)

Una vez se recogen datos, el paso siguiente es crear histogramas para visualizar la forma de las distribuciones.

Considere el conjunto de observaciones x_1, x_2, \dots, x_n

Para crear histogramas con intervalos iguales, se siguen los siguientes pasos.

Paso 1. Calcule el número k de intervalos de igual longitud. Si el número de datos es n , se usa la regla de Sturges para estimar el número de clases: $k = \text{roundup}(\log_2(n)+1)$ Donde $\text{roundup}()$ significa que se redondea la fracción al siguiente entero y $\log_2()$ es la función logaritmo base 2.

Paso 2. Calcule los extremos de los intervalos, $q(i)$ usando $q(i) = \min(x_1, x_2, \dots, x_n) + i * [\max(x_1, x_2, \dots, x_n) - \min(x_1, x_2, \dots, x_n)] / k$ para $i=0, \dots, k$

Paso 3. Cuente el número de observaciones o frecuencia en cada clase, $C(i)$ para $i=1, 2, \dots, k$ y las frecuencias relativas, $rf(i)$ usando

$rf(i) = C(i)/(n)$ para $i=1, 2, \dots, k$

Número de clases recomendadas para un histograma de acuerdo con regla de Sturges

n número de clases = $\text{roundup}(1.442 \ln(n)+1)$

10 5

20 6

50 7

100 8

200 9

1000 11

10000 15

Suponga que se tienen $n=9$ observaciones

```
tiempos=c(7.2, 4.5, 8.1, 9.2, 4.2, 12.3, 15.1, 6.2, 4.8);  
plot(tiempos);
```

De acuerdo con la regla de Sturges, se necesitan 5 clases. Los extremos de los intervalos de clases son $q(0)=4.2$, $q(1)=6.4$, $q(2)=8.6$, $q(3)=10.7$, $q(4)=12.9$, y $q(5)=15.1$. Las frecuencias relativas son $rf(1)=0.4$, $rf(2)=0.2$, $rf(3)=0.1$, $rf(4)=0.1$, y $rf(5)=0.1$

```
rf=c(0.4, 0.2, 0.1, 0.1, 0.1);
```

```
barplot(rf, main="Histograma", xlab="tiempos", ylab="frecuencia relativa",
, names.arg=c("4.2-6.4", "6.4-8.6", "8.6-10.7", "10.7,12.9", "12.9-15.1"));
```

Puede hacer el histograma directamente en R con la función `hist()`. Otros paquetes de R, como `ggplot`, también tienen funciones que elaboran histogramas (ver por ejemplo, <http://www.harding.edu/fmccown/r/>).

Note que el histograma elaborado con R o con excel tiene un número distinto de clases; Excel, por ejemplo, usa menos clases para conjuntos pequeños de datos.

```
hist(tiempos);
```

Note también que `hist()` no calcula frecuencias relativas directamente; para esto, puede buscar otros paquetes, calcularlas a partir de las frecuencias:

```
h <- hist(tiempos, breaks = 6, plot=FALSE)
h$counts=h$counts/sum(h$counts)
plot(h)
```