

Ejercicio resuelto 5.1

Un Centro de Estudios cuenta con 20 aulas, de las cuales 6 tienen 10 puestos, 5 tienen 12 puestos, 4 tienen 15 puestos, 3 tienen 20 puestos y 2 tienen 25 puestos.

- a) ¿Cuántos puestos tiene a lo sumo el 25 % de las aulas más pequeñas?
- b) ¿Cuántos puestos tiene como mínimo el 25 % de las aulas más grandes?

Solución:

El **objetivo del Apartado a)** es determinar cuántos puestos tiene a lo sumo el 25 % de las aulas más pequeñas.

Planteamiento: el experimento consiste en seleccionar aulas (individuos) y observar el número de puestos que tiene (variable). La población son todas las aulas de ese centro y se tiene una muestra (realmente es un censo) de 20 aulas. La variable es cardinal (porque los valores son números), de razón (porque 0 significa que no hay puestos) y discreta, porque no puede haber, por ejemplo, 1,5 puestos.

Método y justificación: se busca un punto que separe el 25 % de los aulas más pequeñas del 75 % de las más grandes.

Gráficamente, se puede plantear el problema como sigue: se representa el número de puestos en un eje ordenando los valores de menor a mayor (de aulas pequeñas a grandes). Se marca la zona que se corresponde al 25 % de las más pequeñas y se llama a al valor más bajo de esa zona y b al más alto. Así, ese 25 % tendrá como mínimo a puestos y como máximo b puestos, por lo que nuestro objetivo será buscar el valor de b (ver Figura 5.1).



Figura 5.1: Eje para buscar el cuantil en el Apartado a).

Es el mismo problema que el de la mediana cambiando el 50 % por el 25 %, es decir, se debe encontrar $b = Q_{0,25}$, el cuantil de orden 0,25.

Cálculos: se utilizará la tabla de frecuencias relativas acumuladas. Como son datos aislados, se busca el valor 0,25 en la columna de frecuencias relativas acumuladas y se aplica el mismo razonamiento que para la mediana:

- Si 0,25 no está en la columna de frecuencias acumuladas, $b = Q_{0,25}$ será el primer valor para el que F_i sobrepasa el 0,25.
- Si 0,25 está en la columna de frecuencias acumuladas, $b = Q_{0,25}$ sería cualquier valor entre el x_i correspondiente y el siguiente. Por ejemplo, se podría tomar el punto medio.

x_i	f_i	F_i
10	6	0,3
12	5	0,55
15	4	0,75
20	3	0,9
25	2	1
Total	20	–

Tabla 5.1: Tabla de frecuencias para el Apartado a).

El valor $F_i = 0,25$ no aparece en la Tabla 5.1, y que el primer valor cuyo F_i sobrepasa 0,25 es $x_i = 10$, luego $b = Q_{0,25} = 10$ puestos.

Conclusión: como $Q_{0,25} = 10$, el 25 % de las aulas más pequeñas de ese Centro de Estudios tienen a lo sumo 10 puestos. En realidad como el número mínimo de puestos también es 10, todas las aulas que forman el 25 % de las más pequeñas tienen exactamente 10 puestos.

Problema propuesto: Apartado b).

Ejercicio resuelto 5.2

Se ha realizado un estudio acerca del tiempo que navegan al día los usuarios de internet. Los datos aparecen en la siguiente tabla:

Tiempo (en minutos)	0-60	60-120	120-180	180-240	240-480
Porcentaje de usuarios	48,98	22,51	10,78	9,02	8,71

- a) Si se considera el 20 % de la gente que más navega, ¿en qué rango de tiempos se podría decir que se encuentran?
- b) ¿Cuánto tiempo navega como máximo el 30 % de los que menos tiempo acceden?

Solución:

El **objetivo del Apartado a)** es determinar en qué rango de tiempos se encuentra el 20 % de la gente que más navega.

Planteamiento: el experimento consiste en seleccionar usuarios de internet (individuos) y observar el tiempo que navegan (variable). La población son todos los usuarios de internet y se desconoce el tamaño de la muestra. La variable es cardinal (por que los valores son números), de razón (porque 0 significa que no navega) y continua (porque podría darse cualquier valor, depende de la precisión). No se conocen los datos aislados, solo algunos rangos con sus frecuencias.

Método y justificación: se busca encontrar el intervalo de tiempo en el que se mueve el 20 % de los que más navegan. En la Figura 5.3 se representa el 20 % de los que más navegan con tiempos comprendidos entre un valor a y un valor b . Como en este caso el valor b es el máximo, por los datos del enunciado se sabe que es 480 minutos. Para hallar el valor a se deben hacer algunos cálculos a partir de las frecuencias.

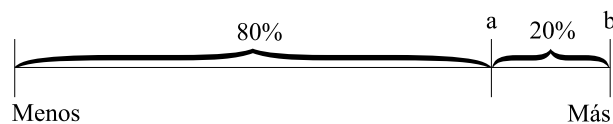


Figura 5.2: Eje para buscar el cuantil en el Apartado a).

Como la tabla de frecuencias está ordenada de menor a mayor, *siempre* se deben hacer los cálculos teniendo en cuenta lo que se acumula por debajo, no por encima. Es decir, como se observa en la Figura 5.3, se debe hallar el valor a que deja el 80 % por debajo, es decir, $a = Q_{0,8}$.

Cálculos: al igual que se hacía para calcular la mediana, se busca el intervalo en el que se sobrepasa por primera vez el 80 % de frecuencia acumulada en la Tabla 5.2. El 71,49 % navega 120 minutos o menos, mientras que el 82,27 % navega como mucho 180 minutos, luego se pasa del 80 % justo en el intervalo (120, 180].

Clases	f_i	F_i
[0, 60]	0,4898	0,4898
(60, 120]	0,2251	0,7149
(120, 180]	0,1078	0,8227
(180, 240]	0,0902	0,9129
(240, 480]	0,0871	1
Total	1	—

Tabla 5.2: Tabla de frecuencias para Apartado a).

Para aproximar le valor de $Q_{0,8}$, se puede interpolar suponiendo que los tiempos están repartidos más o menos uniformemente en ese intervalo.

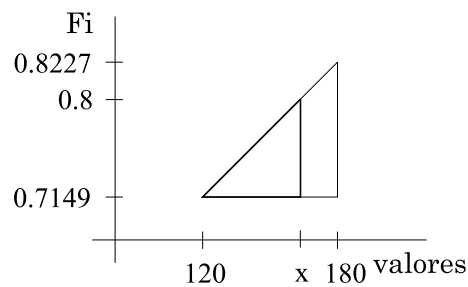


Figura 5.3: Interpolación para el cuantil en el Apartado a).

Aplicando triángulos semejantes según se muestra en la Figura 5.3 (base grande es a altura grande lo mismo que base pequeña es a altura pequeña), se tiene que:

$$\begin{array}{rcl} 180 - 120 = 60 & \text{—} & 0,8227 - 0,7149 = 0,1078 \\ x - 120 & \text{—} & 0,8 - 0,7149 = 0,0851 \end{array}$$

En consecuencia $x - 120 = 0,0851 \times 60 / 0,1078$, es decir $x = 47,3655 + 120 = 167,3655$ minutos.

Conclusión: el 80 % de los usuarios que menos navegan, están a lo sumo unos 167,3655 minutos conectados. Esto es una aproximación porque al interpolar se supone que había uniformidad en el intervalo y no tiene por

qué ser exacta. En conclusión, se puede decir que el 20 % de los internautas de la muestra que más navegan están entre 167,3655 y 480 minutos al día conectados.

Problema propuesto: Apartado b).

Ejercicio resuelto 5.3

Según datos de un estudio de Turismo urbano, en cierta localidad hay 3 establecimientos hoteleros con tarifa de 60€ por noche, 5 con tarifa de 50€, 4 con tarifa de 45€, 5 con tarifa de 36€, 8 con tarifa de 30€ y 5 con tarifa de 25€. Además han verificado que 12 de esos establecimientos tienen menos de 15 plazas, 8 tienen entre 15 y 30 plazas y el resto tienen entre 30 y 70 plazas.

- a) ¿Cuántos hoteles tienen una tarifa superior a 40€?
- b) ¿Qué porcentaje de establecimientos tienen una tarifa comprendida entre los 35 y los 50€?
- c) ¿Cuántos hoteles tienen aproximadamente entre 20 y 30 plazas?, ¿qué porcentaje representan respecto al total?

Solución:

El **objetivo del Apartado a)** es localizar cuántos hoteles tienen una tarifa superior a 40€.

Planteamiento: el experimento consiste en seleccionar los establecimientos hoteleros de ese lugar (individuos) y observar su tarifa (variable 1) y el número de plazas (variable 2). La población la constituyen todos los establecimientos hoteleros y se tiene una muestra (que es realmente un censo) de 30. Las variables son cardinales (porque los valores son números), de razón (porque 0 significaría que no cuesta nada para la variable 1 y que no tiene ninguna plaza para la variable 2, es decir, representa la nulidad). La variable 1 está actuando como variable discreta al tomar pocos valores distintos y la variable 2 actúa como continua. El enunciado proporciona datos aislados para la variable 1 y datos agrupados para la variable 2. El Apartado a) concierne sólo a la variable 1.

Método y justificación: se debe calcular el número de hoteles que cuestan más de 40€. Con la notación de la Figura 5.4 por debajo de una tarifa de 40€ habrá un número a de hoteles y por encima un número b , luego se tiene que calcular b .

Como la tabla de frecuencias se ordena de menor a mayor, conviene hacer los cálculos a partir de lo que se acumula por debajo, es decir, se buscará primero a a partir de la tabla de frecuencias y b será lo que quede hasta sumar los 30 hoteles. El valor de a indica la *posición absoluta* que ocupa el valor 40 dentro de la muestra.

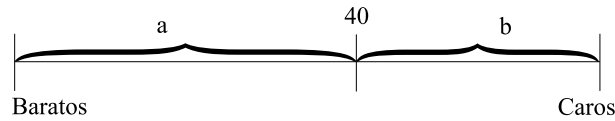


Figura 5.4: Eje para la posición absoluta en el Apartado a).

Cálculos: como se necesita el número de hoteles y no el porcentaje, se busca en las frecuencias acumuladas absolutas. Según la Tabla 5.3 el valor 40 está entre el 36 y el 45, luego por debajo de 40 hay 18 hoteles y por encima estaría el resto: $30 - 18 = 12$.

x_i	n_i	N_i
25	5	5
30	8	13
36	5	18
45	4	22
50	5	27
60	3	30
Total	30	–

Tabla 5.3: Tabla de frecuencias para el Apartado a).

Conclusión: 12 de los 30 hoteles de la muestra ofrecen una tarifa superior a 40€.

El **Apartado b)** también concierne a la variable 1. El **objetivo** es localizar el porcentaje de hoteles con tarifa comprendida entre los 35 y los 50€.

Método y justificación: con la notación de la Figura 5.5 entre 35 y 50€ hay un porcentaje c de hoteles. Como la tabla de frecuencias se ordena de menor a mayor, conviene hacer los cálculos a partir de lo que se acumula por debajo. En este caso, se calcularía en primer lugar lo que se acumula por debajo de 35 y luego lo que se acumula por debajo de 50. Así, lo que se acumula entre 35 y 50 será la diferencia entre esas dos cantidades, es decir, $c = b - a$. El valor de a indicaría la *posición relativa* que ocupa el valor 35 dentro de la muestra. Como se necesitan porcentajes, se calcularán a partir las frecuencias acumuladas relativas.

Cálculos: según la Tabla 5.4 el valor 35 está entre el 30 y el 36, luego por debajo de los 35€ se encuentra el 43,33 % de los establecimientos hoteleros. Además la tarifa de 50€ aparece en la tabla con una frecuencia

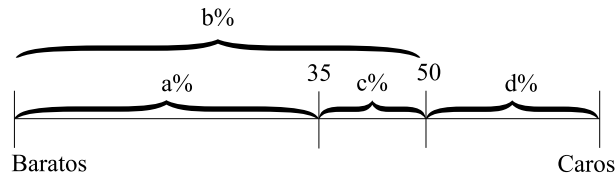


Figura 5.5: Eje para la posición relativa en el Apartado b).

acumulada del 90 %. Como por debajo de 50 está el 90 % y por debajo de 35 está el 43,33 %, entre 35 y 50 estará el $90 - 43,33 = 46,47$ %.

x_i	n_i	N_i	F_i
25	5	5	0,1667
30	8	13	0,4333
36	5	18	0,6
45	4	22	0,7333
50	5	27	0,9
60	3	30	1
Total	30	—	—

Tabla 5.4: Tabla de frecuencias para el Apartado b).

Conclusión: el 46,47 % de los establecimientos hoteleros ofrecen tarifas comprendidas entre los 35 y los 50€.

El **Apartado c)** es similar al Apartado b), pero se refiere a la variable 2, para la que se tienen los datos agrupados. Al igual que se hacía en el cálculo de los cuantiles, en este caso se aproximan los valores que se necesitan interpolando. El **objetivo** es calcular el número aproximado de hoteles que tienen entre 20 y 30 plazas e indicar el porcentaje que representan del total.

Método y justificación: con la notación de la Figura 5.6 hay un número c de hoteles que tienen entre 20 y 30 plazas. Como siempre, conviene hacer los cálculos a partir de lo acumulado por debajo de cada valor, por lo que se calculará primero el número a de hoteles con menos de 20 plazas, luego el número b de hoteles con menos de 30 plazas y así c será la diferencia entre ambos números.

Cálculos: como se necesitan números absolutos y no porcentajes, se busca en las frecuencias acumuladas absolutas. Según la Tabla 5.5 el valor 20 está en el intervalo de 15 a 30 plazas. Hay acumulados 12 hoteles de hasta

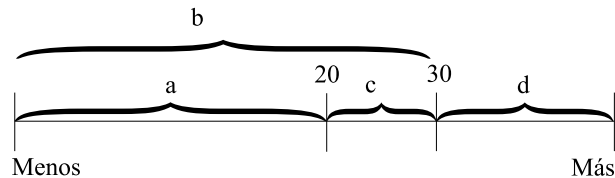


Figura 5.6: Eje para la posición relativa en el Apartado c).

15 plazas y que hasta 30 plazas se acumulan un total de 20. Para saber cuántos se acumulan aproximadamente hasta 20, se puede interpolar.

Clases	n_i	N_i
$(0, 15]$	12	12
$(15, 30]$	8	20
$(30, 70]$	10	30
Total	30	—

Tabla 5.5: Tabla de frecuencias para el Apartado c).

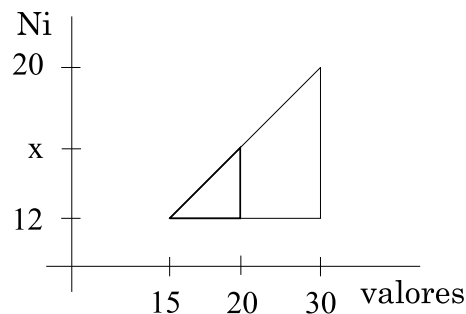


Figura 5.7: Interpolación - posición absoluta en el Apartado c).

Aplicando triángulos semejantes según se muestra en la Figura 5.7 (base grande es a altura grande lo mismo que base pequeña es a altura pequeña), se tiene que:

$$\begin{array}{rcl} 30 - 15 = 15 & \text{——} & 20 - 12 = 8 \\ 20 - 15 = 5 & \text{——} & x - 12 \end{array}$$

En consecuencia $x - 12 = 5 \times 8 / 15$, es decir $x = 2,6667 + 12 = 14,6667$ establecimientos hoteleros. Hay aproximadamente unos 14,7 establecimientos hoteleros que tienen 20 plazas o menos.

Ahora se debe calcular cuántos establecimientos tienen 30 plazas o menos para luego hacer la diferencia. En la Tabla 5.5 aparece el valor de 30 plazas. Hasta ese punto se acumulan exactamente 20 establecimientos. En conclusión, si hay 20 con 30 plazas o menos y hay unos 14,7 con 20 plazas o menos, entre 20 y 30 plazas habrá $20 - 14,7 = 5,3$.

Conclusión: hay aproximadamente 5,3 establecimientos que tienen entre 20 y 30 plazas. Para saber qué porcentaje representan, basta tener en cuenta que hay un total de 30 establecimientos. Como $5,3/30 = 0,1767$, se concluye que aproximadamente el 17,67% de los establecimientos hoteleros de ese lugar tienen entre 20 y 30 plazas.

Ejercicio resuelto 5.4

Un comercial opera con dos empresas de las que percibe semanalmente alrededor de 165,66 y 216,87€ de comisión respectivamente. El importe de las comisiones que paga la primera empresa a sus agentes sigue una distribución de media 114,46€/semana y desviación de 37,35€/semana. La segunda empresa presenta una comisión media de 191,57€/semana y una desviación de 56,63€/semana. ¿En cuál de las dos empresas ocupa mejor posición relativa aquel agente comercial?

Solución:

El **objetivo** es determinar en qué empresa ocupa mejor posición relativa el agente comercial.

Planteamiento: el experimento consiste en seleccionar comerciales de la Empresa 1 y de la Empresa 2 (individuos) y observar la comisión que les pagan. Se tienen dos poblaciones, por un lado se tienen los agentes de la Empresa 1 (Población 1), y por otro los de la Empresa 2 (Población 2). Las muestras, de tamaño desconocido, coinciden con las poblaciones (aunque el enunciado sólo proporciona las medias y las desviaciones típicas). Las variables son cardinales, de razón y continuas porque los valores son números, el 0 significa ausencia de comisión y, en principio, cualquier valor es válido.

Método y justificación: se debe comparar la posición que ocupa el comercial en cada empresa relativamente (es decir, en comparación con el resto de los agentes de su empresa). No se deben comparar las comisiones de las dos empresas en términos absolutos: 165,66€ frente a 216,87€, porque esas cantidades sean grandes o pequeñas dependen de lo que ganen los demás agentes. Por ello, se debe relativizar.

En principio, se podría resolver el problema tanto con el rango cuantil (ya que el porcentaje acumulado indicaría el lugar que ocupa en cada empresa), como la tipificación, aunque en este caso, como no hay datos para calcular el rango cuantil, sólo se puede utilizar la tipificación.

Para que la tipificación sea realmente fiable, la media debe ser moderadamente representativa. Para ver si esto es cierto, como se trata de variables positivas con la media alejada de 0, se puede calcular el CV.

Cálculos: para distinguir las variables, se llamará X a la variable de la Población 1 e Y a la variable de la Población 2. Se tienen las medias y

las desviaciones calculadas, luego sólo hace falta aplicar las fórmulas.

$$CV_x = \frac{S_x}{|\bar{x}|} = \frac{37,35}{114,46} = 0,3263 \quad \text{y} \quad CV_y = \frac{S_y}{|\bar{y}|} = \frac{56,63}{191,57} = 0,2953.$$

$$z_{x_i} = \frac{x_i - \bar{x}}{S_x} = \frac{165,66 - 114,46}{37,35} = 1,3708,$$
$$z_{y_i} = \frac{y_i - \bar{y}}{S_y} = \frac{216,87 - 191,57}{56,63} = 0,4458.$$

Conclusión: como el CV de las comisiones se sitúa alrededor del 30 % en ambas empresas, las medias son bastante representativas y se pueden establecer de manera fiable las comparaciones relativas mediante la tipificación.

El valor tipificado del agente para la primera empresa es de 1,3708 y para la segunda de 0,4458. En ambos casos el comercial está por encima de la media, ya que la tipificación da valores positivos. Sin embargo, la puntuación relativa es mayor en la primera empresa que en la segunda (es decir, está “más por encima de la media” en la primera empresa), luego ocupa una mejor posición relativa en esa primera empresa, a pesar de cobrar más en la segunda.

Los valores de la tipificación son “normales” (menores que 2). El valor de 0,4458 indica que el comercial está por encima de la media, pero no mucho, porque es un valor bastante cercano a 0, mientras que en la segunda sí hay más diferencia, aunque tampoco se trata de valor extremo, ya que está completamente dentro del rango de los valores habituales de la tipificación.