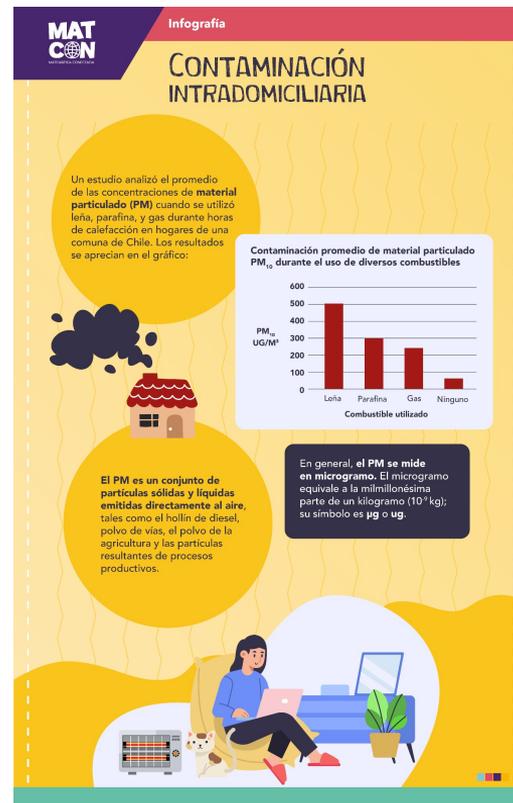




Calefaccionando el hogar

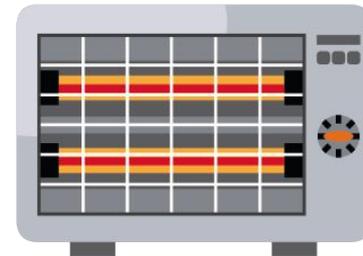
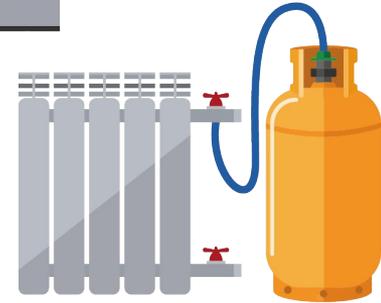


Infografía: Contaminación intradomiciliaria



Contaminación intradomiciliaria

1. ¿Qué combustible genera mayores concentraciones promedio de materiales particulados en el aire?
2. Si el gas genera menores concentraciones de promedio de materiales particulados en el aire, ¿por qué la parafina o la leña se usan con mayor frecuencia?



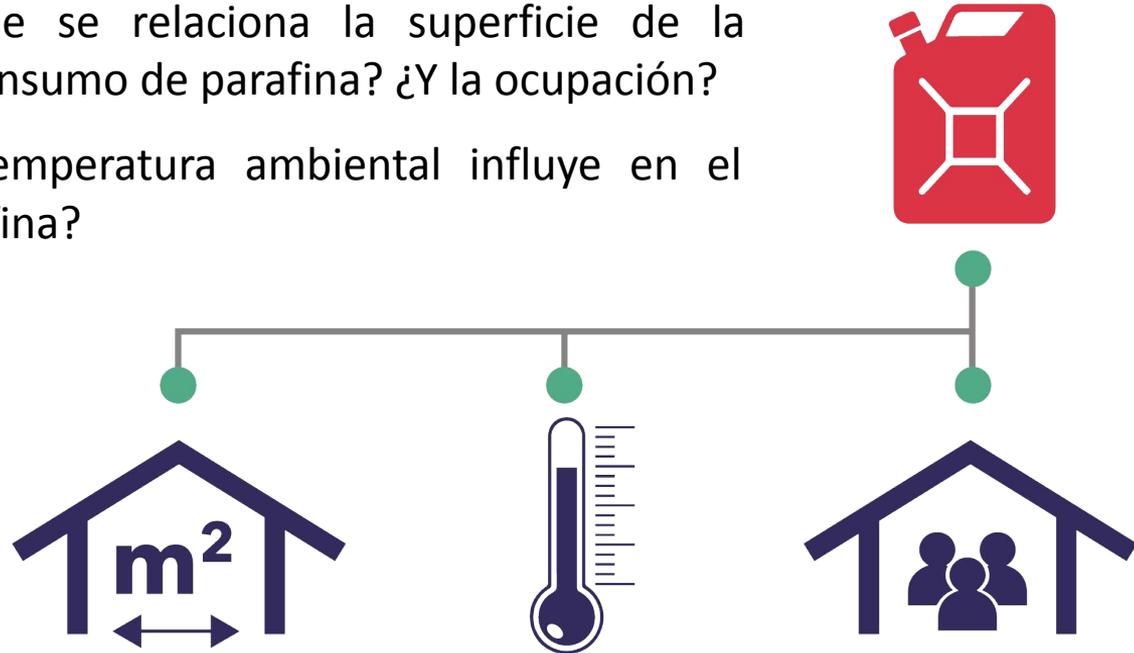
Presentación del problema

Se encargó un estudio para evaluar el consumo de parafina en los meses de invierno en una ciudad del sur de Chile. Para ello, se eligió una muestra de 30 hogares, y se recopilaron datos de las siguientes variables:

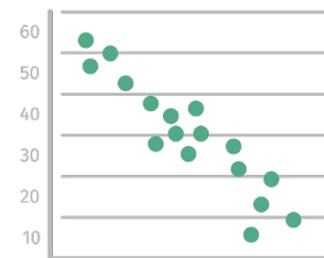
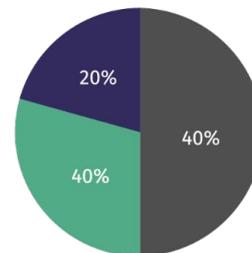
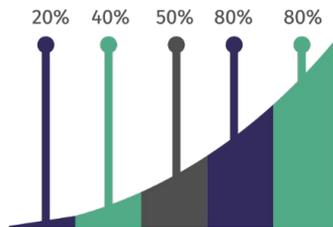
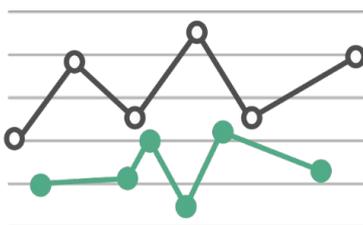
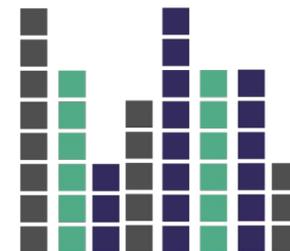
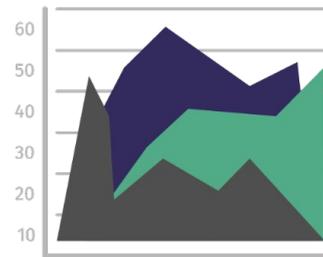
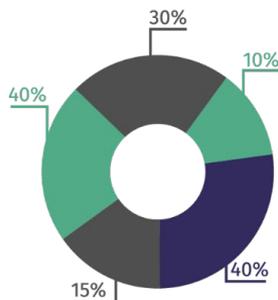
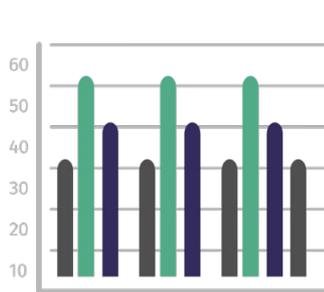
Consumo semanal de parafina	Cantidad de litros de parafina consumidos durante la semana
Superficie de la vivienda	Cantidad de metros cuadrados construidos de la vivienda
Temperatura promedio semanal	Promedio semanal de las temperaturas medias diarias
Ocupación promedio de la vivienda	Promedio semanal de horas en la que hubo al menos una persona en la vivienda durante el día

Actividad 1

1. ¿Cómo crees que se relaciona la superficie de la vivienda con el consumo de parafina? ¿Y la ocupación?
2. ¿Crees que la temperatura ambiental influye en el consumo de parafina?

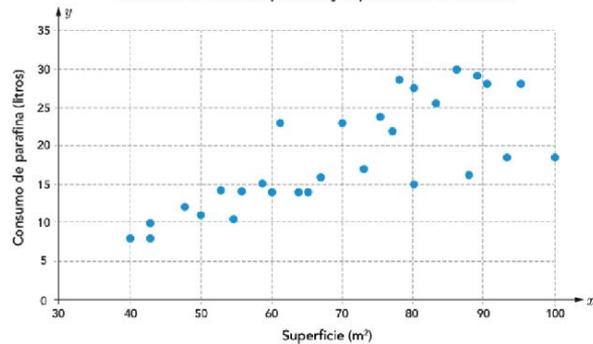


¿Qué tipo de gráfico es el más adecuado para analizar la relación entre dos variables cuantitativas?

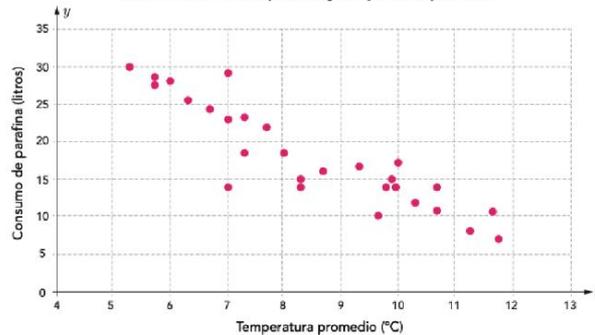


Actividad 2

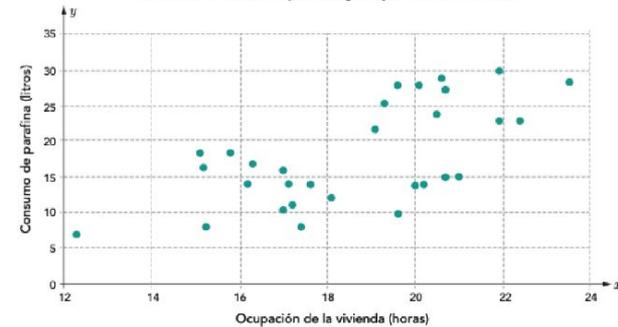
Consumo semanal de parafina y superficie de la vivienda



Consumo semanal de parafina y temperatura promedio

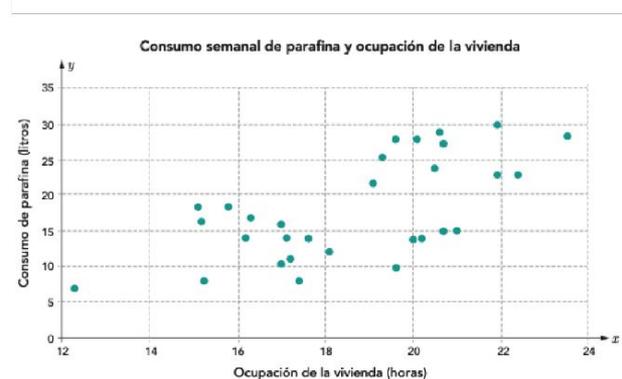
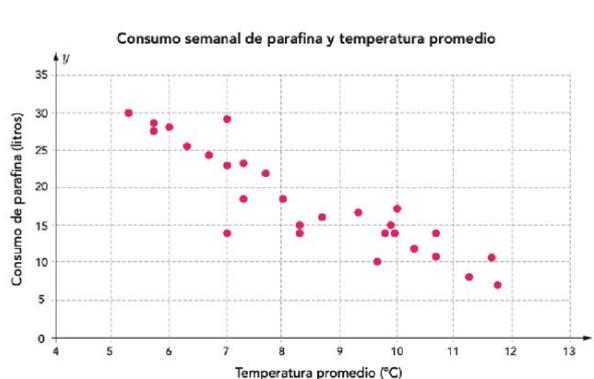
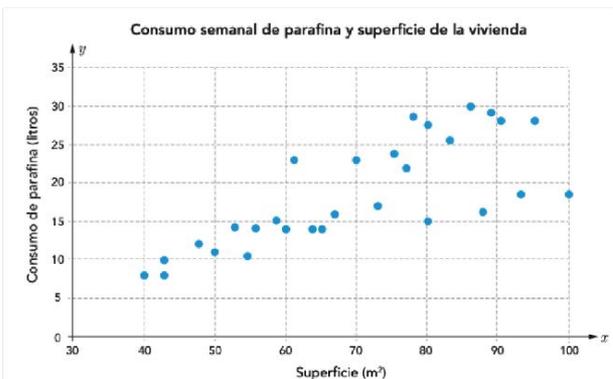


Consumo semanal de parafina y ocupación de la vivienda



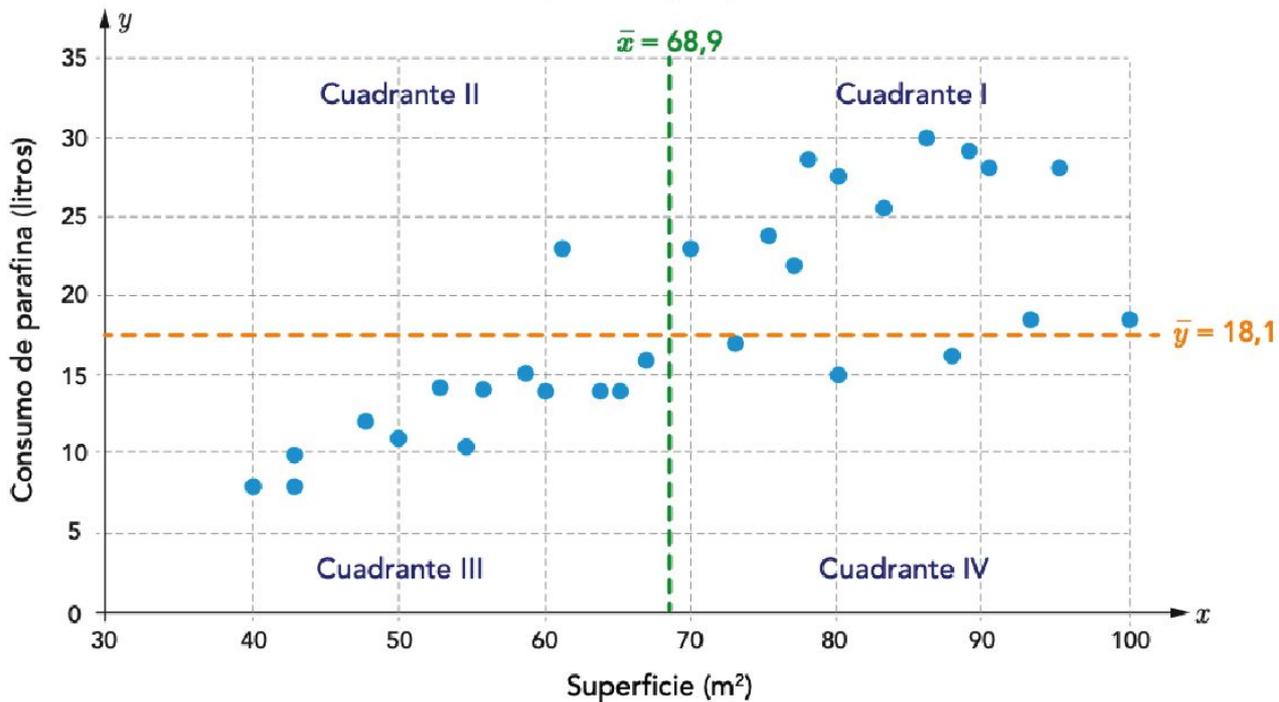
Actividad 2

1. Considerando los gráficos, ¿se puede establecer cuál de las variables está más relacionada con el consumo semanal de parafina? ¿Por qué?



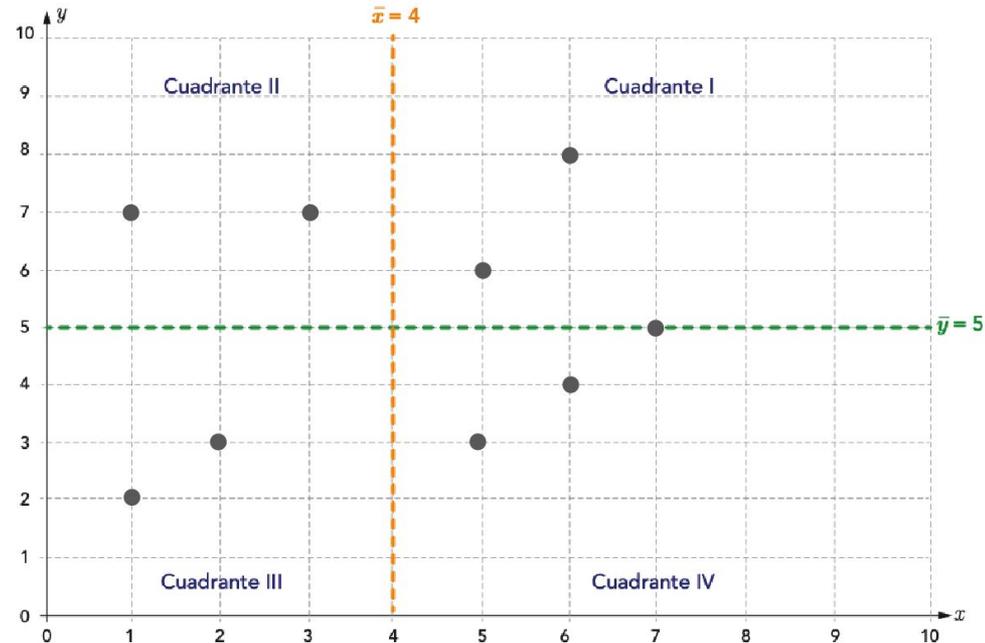
Actividad 3

Consumo semanal de parafina y superficie de la vivienda



Actividad 3

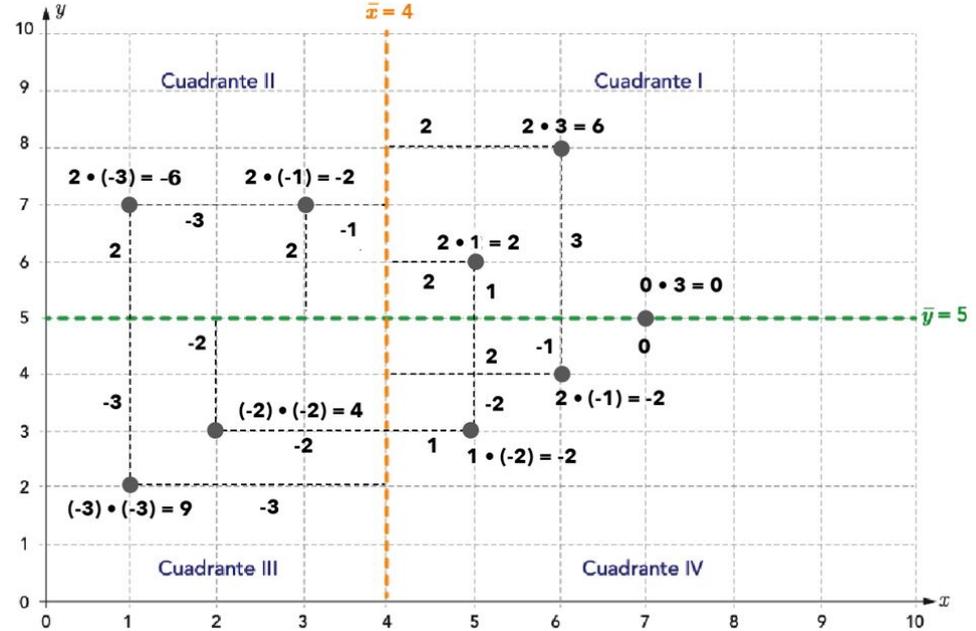
- ¿Qué se puede decir de la ubicación de los puntos en los cuadrantes?
- ¿Qué pasa con el producto de las diferencias respecto a las medias?



Actividad 3

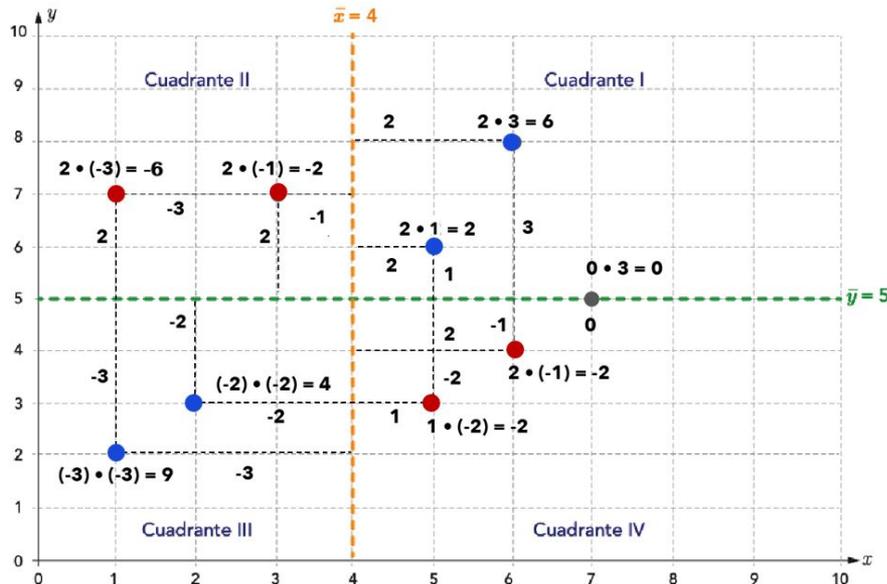
Respecto a la variable x , la diferencia de cada variable con respecto a su media será: positiva en el I y IV cuadrante y negativa en el II y III cuadrante.

Respecto a la variable y , la diferencia de cada variable con respecto a su media será: positiva en el I y II cuadrante y negativa en el III y IV cuadrante.



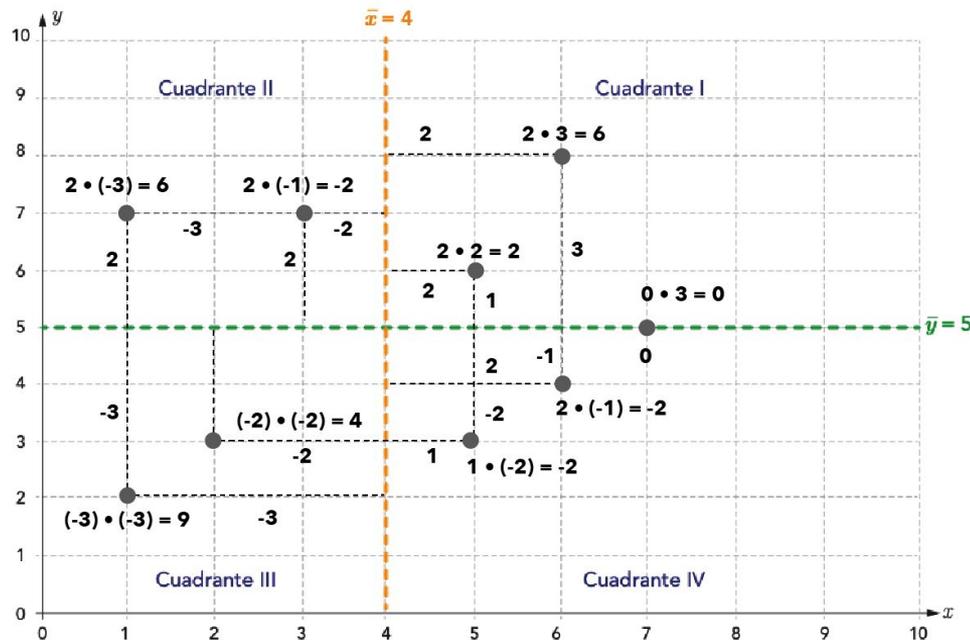
Actividad 3

Por lo tanto, el producto de estas diferencias es siempre positivo para los puntos ubicados en los cuadrantes I y III, mientras que es negativo para los puntos ubicados en los cuadrantes II y IV.



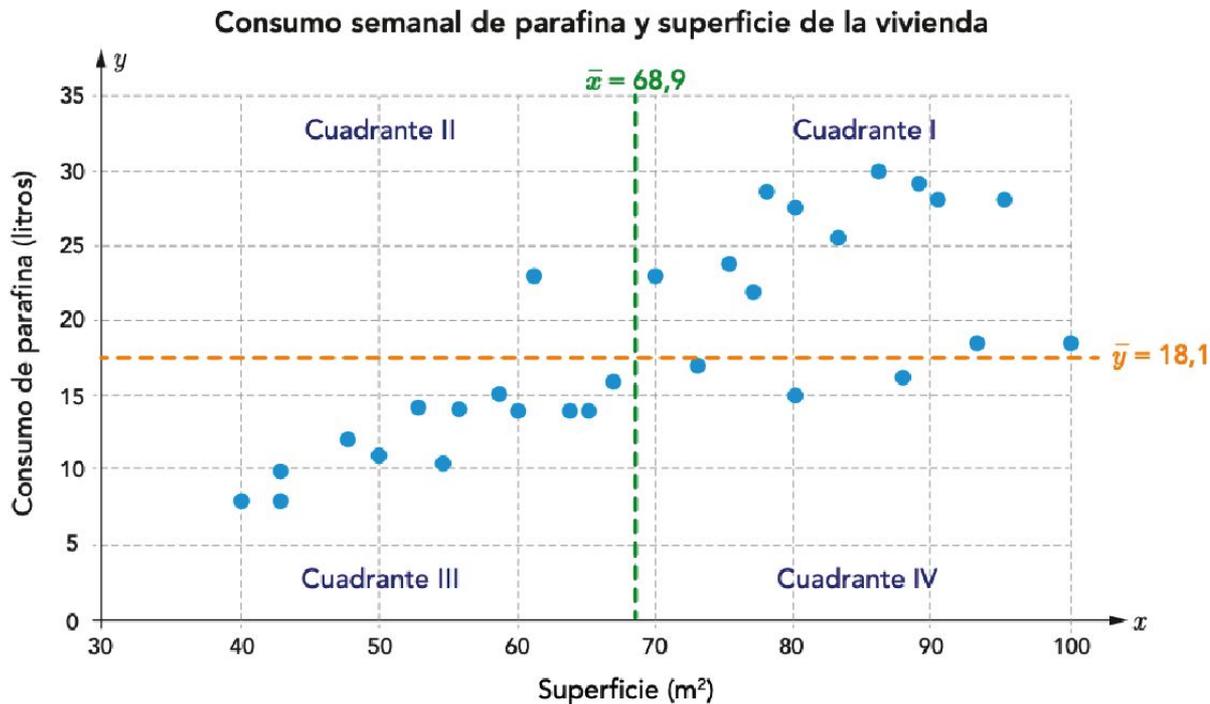
Actividad 3

El producto de estas diferencias es siempre positivo para los puntos ubicados en los cuadrantes I y III, mientras que es negativo para los puntos ubicados en los cuadrantes II y IV.



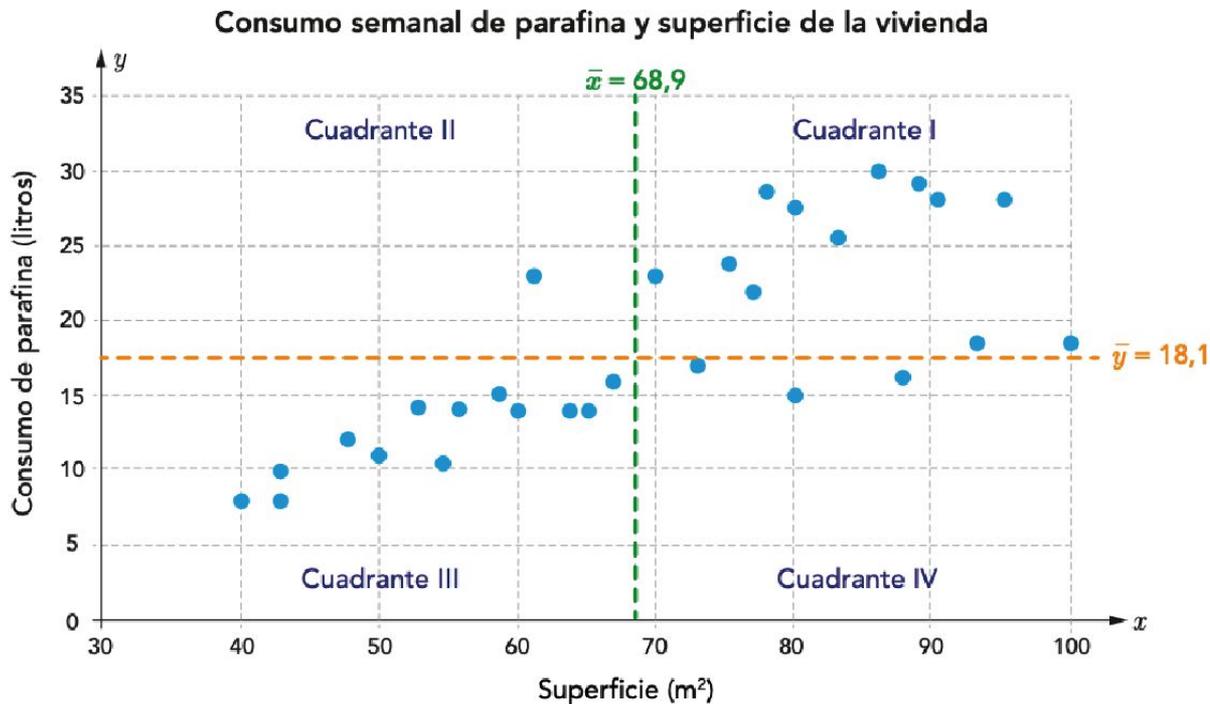
Actividad 3

Solamente mirando el gráfico, ¿hay más puntos cuyos productos de sus diferencias respecto a las medias muestrales son positivos o negativos?



Actividad 3

¿Cómo se podría encontrar una **medida cuantitativa** que permita dar cuenta de qué tipo de puntos son los que predominan en el gráfico?



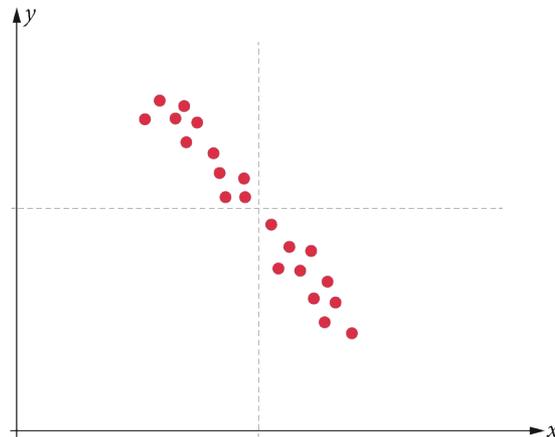
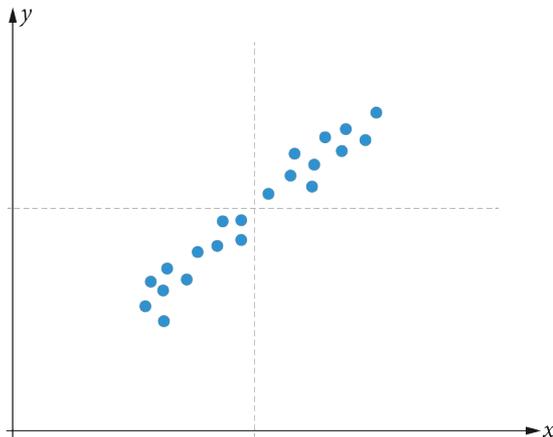
Actividad 3

¿Qué representa esta suma?

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Actividad 3

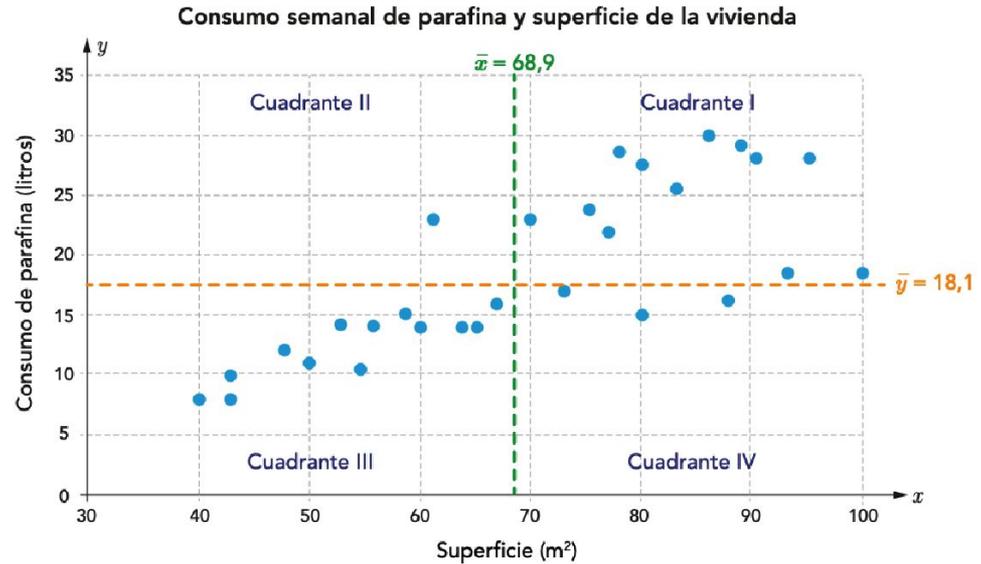
- ¿Qué valores tomaría la suma cuando hay muchos más puntos azules? ¿Y cuándo hay muchos más puntos rojos?
- ¿Es correcto afirmar que siempre que hay más puntos azules que rojos, esta suma será positiva? ¿Qué puede afectar al valor de la suma?



Actividad 3

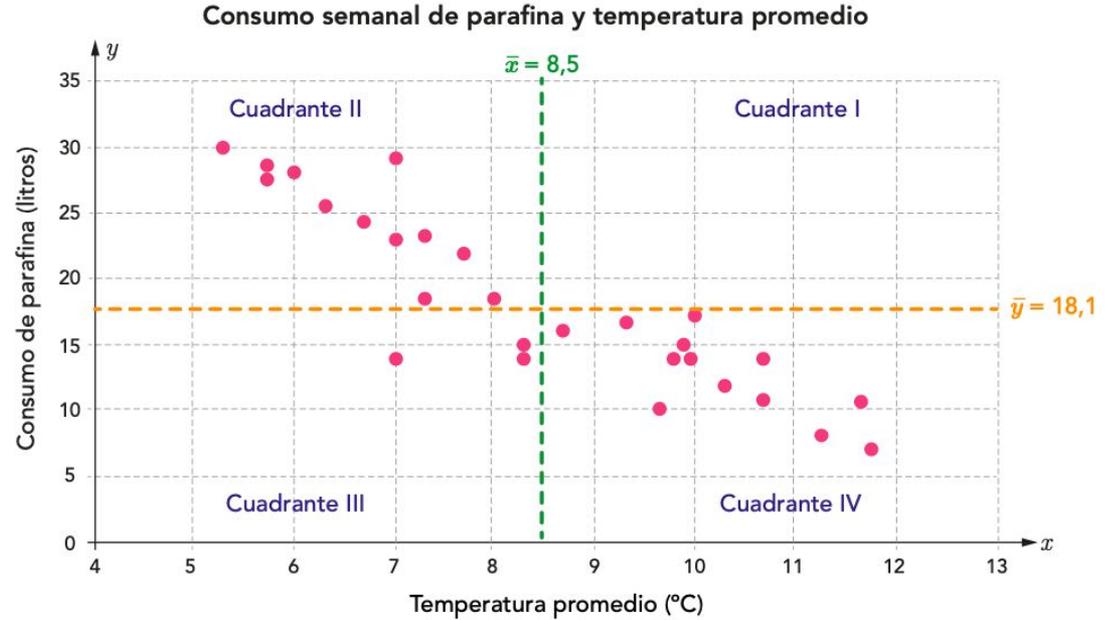
Para este gráfico, ¿la suma de los productos de las diferencias respecto a la media debería ser positiva o negativa?

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 2801,2$$



Actividad 3

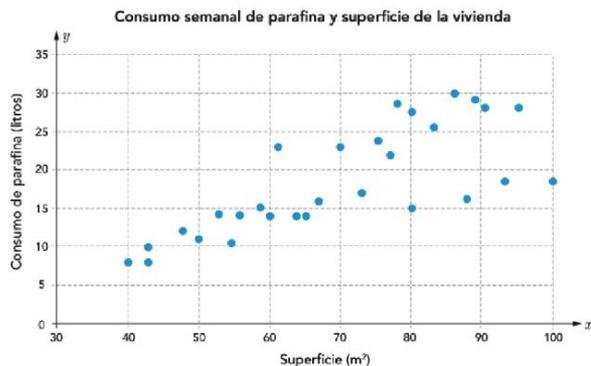
Para este gráfico, ¿la suma de los productos de las diferencias respecto a la media debería ser positiva o negativa?



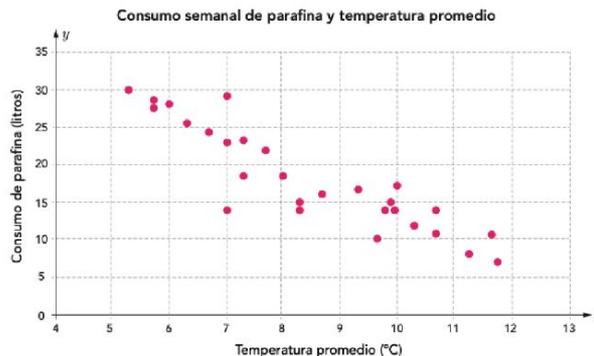
Actividad 4

Variables	Suma de productos
x = superficie del hogar y = consumo semanal de parafina	2801,2
x = temperatura promedio y = consumo semanal de parafina	– 377,74
x = ocupación de la vivienda y = consumo semanal de parafina	345,77

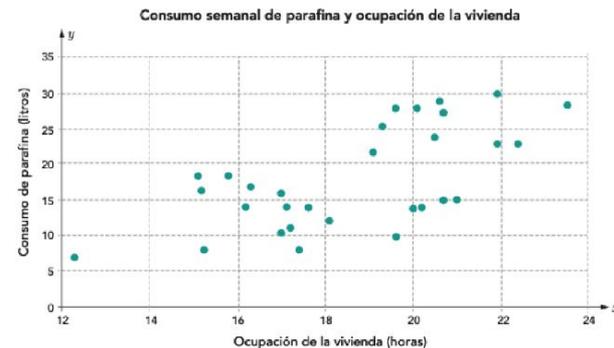
Observando los gráficos y la suma de productos para los tres casos:
¿En cuál de ellos pareciera que los puntos se ven más alineados?



2801,2

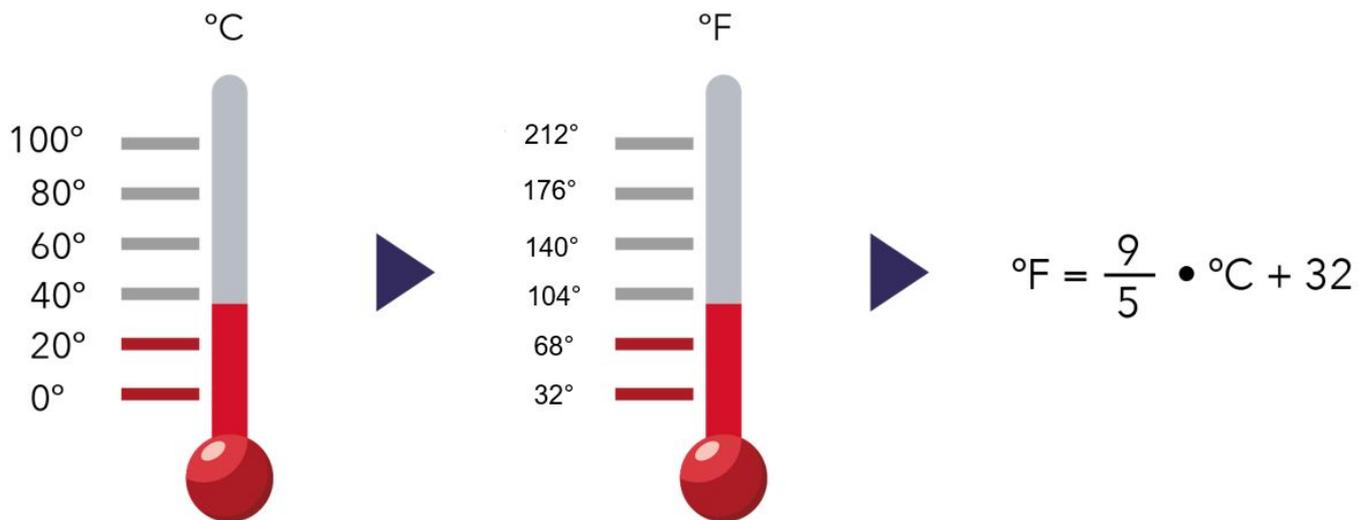


- 377,74



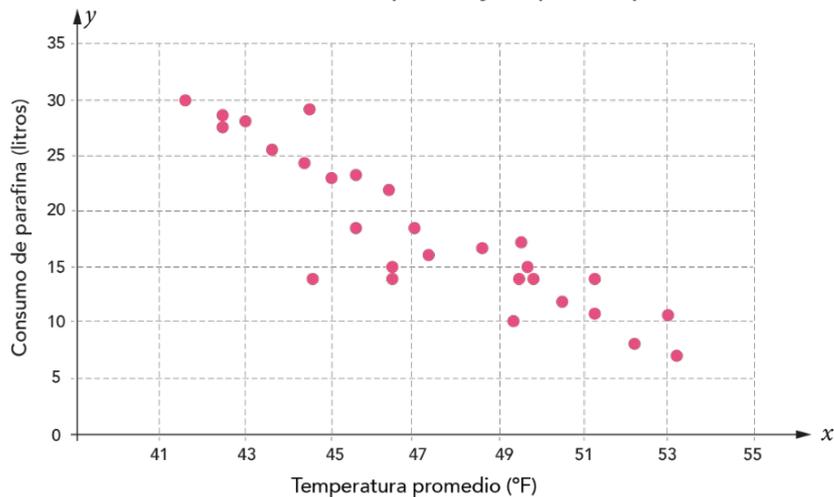
345,77

Considera las variables consumo de parafina y temperatura. Si la temperatura se calcula en grados Fahrenheit en vez de Celsius, ¿la suma de los productos de las diferencias será mayor, menor o igual?

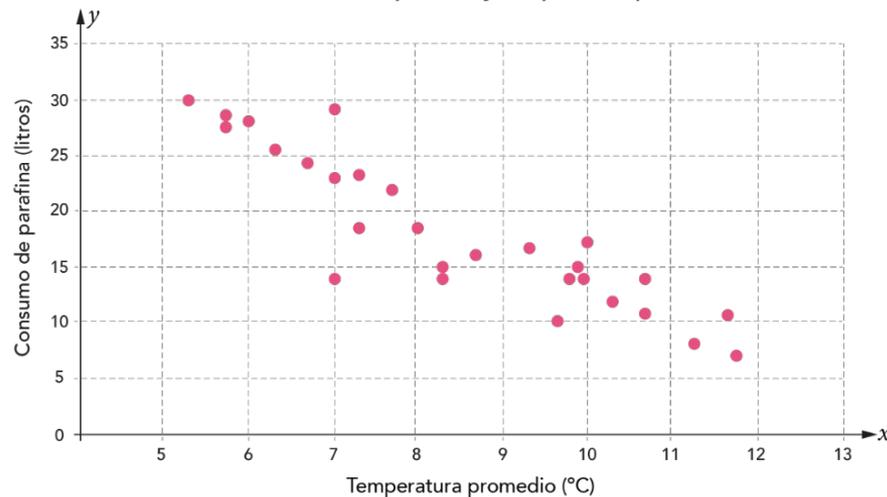


Considera las variables consumo de parafina y temperatura. Si la temperatura se calcula en grados Fahrenheit en vez de Celsius, ¿la suma de los productos de las diferencias será mayor, menor o igual?

Consumo semanal de parafina y temperatura promedio



Consumo semanal de parafina y temperatura promedio



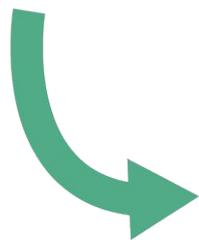
Cambio de escala de medición

Temperatura en °C

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Temperatura en °F

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$



$$\frac{9}{5} \cdot ^\circ\text{C} + 32 = ^\circ\text{F}$$

Sí afecta la suma



No afecta la suma

Cambio de escala de medición

¿Qué efecto tiene?

- No permite usar la suma de productos para **evaluar el grado de asociación**, dada la dependencia en la escala.
- No permite **comparar el grado de asociación** entre pares distintos de variables, dado que éstas posiblemente están medidas en escalas distintas.

Coefficiente de correlación

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

- Siempre es una cantidad entre -1 y 1 .
- Cuando r es muy cercano a -1 o 1 , se interpreta que hay una fuerte relación lineal entre las variables.

En base a estos valores y lo revisado durante la clase...
¿cuál es la variable más relacionada con el consumo semanal de parafina?

Variables	Coefficiente de correlación
x = superficie del hogar y = consumo semanal de parafina	0,782
x = temperatura promedio y = consumo semanal de parafina	- 0,915
x = ocupación de la vivienda y = consumo semanal de parafina	0,644

Conclusiones

- El coeficiente de correlación es una medida estadística que nos indica la **fuerza de la relación lineal entre dos variables**, o conjuntos de datos. Se calcula mediante la siguiente fórmula:

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Esta medida es una versión “estandarizada” de la suma de productos, que no se ve afectada por los cambios de escala.

Conclusiones

El coeficiente de correlación r :

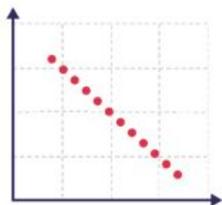
- Es siempre una cantidad **entre -1 y 1** .
- Solo es igual a -1 o 1 cuando entre las variables hay una relación lineal de la forma $y = mx + n$.

Conclusiones

El coeficiente de correlación r :

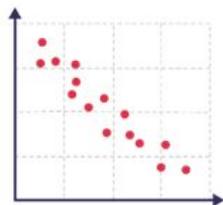
- Cuando r es **positivo**, las variables **crecen o decrecen conjuntamente**
- Cuando r es **negativo**, una **variable decrece a medida que la otra crece**.
- Cuando r es **muy cercano a -1 o 1** , se interpreta que hay una **fuerte relación lineal** entre las variables.

**Correlación
negativa
perfecta**



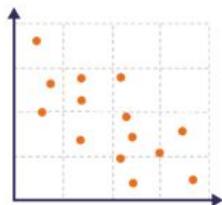
-1

**Correlación
negativa
alta**



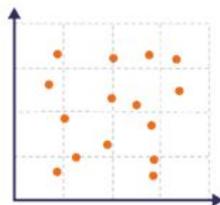
-0.9

**Correlación
negativa
baja**



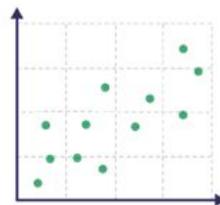
-0.5

**Ausencia de
correlación**



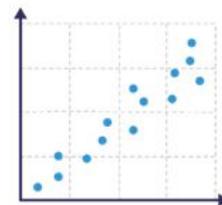
0

**Correlación
positiva
baja**



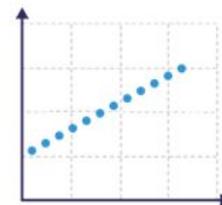
0.5

**Correlación
positiva
alta**



0.9

**Correlación
positiva
perfecta**



1



Calefaccionando el hogar

