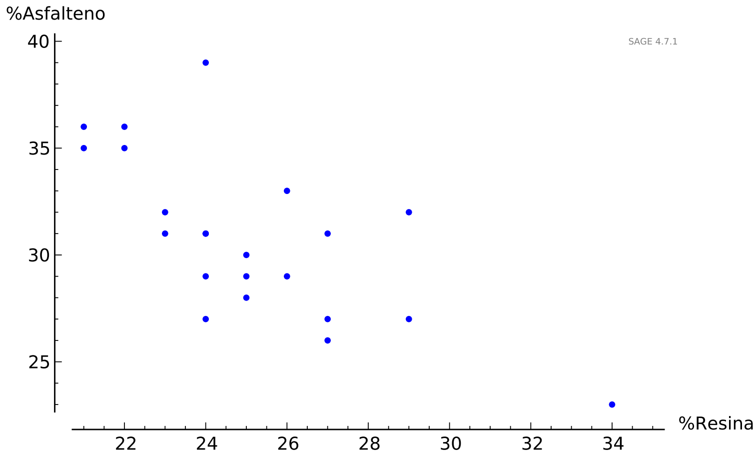
Se quiere estudiar la posible relación lineal entre Y = «Porcentaje de asfalteno» y X = «Porcentaje de resina» en asfaltos utilizados para la fabricación de telas asfálticas. Se dispone de datos de 22 tipos diferentes de asfaltos:

X	22	21	25	23	29	26	25	27	25	21	24	26	23	24	22	27	29	24	24	27	24	34
Y	35	35	29	32	27	29	28	31	30	36	39	33	31	31	36	26	32	31	29	27	27	23

a. Plantear modelo e hipótesis. Mediante el análisis de los residuos, ¿qué se puede decir sobre dichas hipótesis?

Modelo de regresión lineal simple: $Y = \beta_0 + \beta_1 X + U,\ U \sim N(0, \sigma)$.
El análisis de residuos se puede hacer una vez estimada la recta de regresión; entre tanto vemos el diagrama de dispersión de los datos. Se observa una nube de puntos inclinada con pendiente negativa



b. Obtener la recta de regresión y el coeficiente de correlación lineal r. ¿Qué indica el valor del coeficiente de correlación obtenido?

0. $n = 22$
1. $\bar{x} = 25{,}09$ $\qquad \bar{y} = 30{,}77$
2. $n \cdot v_x = (n-1)s_x^2 = 189{,}82$ $\qquad n \cdot v_y = (n-1)s_y^2 = 309{,}86$
3. $n \cdot \text{cov}(x, y) = -168{,}55$ $\qquad r = -0{,}69$
4. $\hat{\beta}_1 = -0{,}89$ $\qquad \hat{\beta}_0 = 53{,}05$
5. $S_R^2 = 8{,}01$
6. Ecuación de la recta de regresión:

$$Y = 53{,}05 - 0{,}89X$$

El coeficiente de correlación lineal $r = -0{,}69$ indica una realción relativamente fuerte entre las variables (el coeficiente de determinación $R^2 = r^2 = 0{,}48$ puede interpretarse como que el modelo explica un $48\,\%$ de la variabilidad de la Y). Su signo negativo indica una relación inversa (cuando X crece Y decrece).

c. ¿Influye el porcentaje de resina sobre el porcentaje de asfalteno? Obtener una conclusión al nivel de significación 0,01.

Estadístico $F = (n-2)\frac{r^2}{1-r^2} = 18{,}68$. Valor crítico: $F_{1,20;0{,}01} = 8{,}017$. Se rechaza la hipótesis nula: $\mathrm{H}_0 \equiv \beta_1 = 0$; por tanto se afirma que el porcentaje de resina (X) influye en el porcentaje de asfalteno (Y).

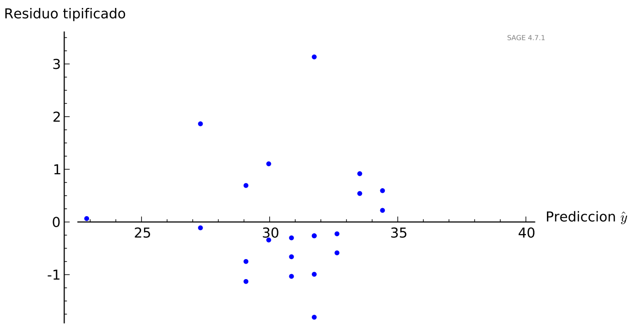
d. Estimar, con una confianza del $95\,\%$, el valor medio del porcentaje de asfalteno para aquellos asfaltos que tienen un $30\,\%$ de resina.

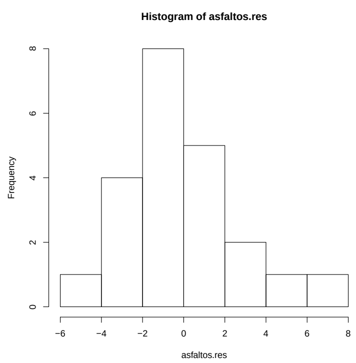
El intervalo estará centrado en la estimación puntual de la media: $\bar{Y}|_{X=30} = 26{,}41$. Nos hace falta el valor crítico $t_{20,0{,}025} = 2{,}086$ y el error típico en la estimación puntual:

$$\sqrt{S_R^2\left(\frac{1}{n} + \frac{(X-\bar{x})^2}{nv_x}\right)} = 1{,}19$$

$$\mathrm{IC}_{95\,\%}(\bar{Y}|_{X=30}) = (26{,}41 \pm 2{,}086 \cdot 1{,}19) = (23{,}93\ ,\ 28{,}90)$$

Una vez estimada la recta de regresión se hallan las predicciones y los residuos. Tipificados estos últimos se obtienen el diagrama de dispersión (contra las predicciones) y el histograma:





A la vista de estos diagramas, no parece haber razones para rechazar normalidad (histograma) u homocedasticidad (diagrama de dispersión).