Se desea estudiar la esperanza de vida `Y` en una serie de países como función de la tasa de natalidad `nat`, la tasa de mortalidad infantil `mortinf` y el logaritmo del producto nacional bruto `lpnb`. Para ajustar el modelo

$$Y_i = \beta_0 + \beta_1 \texttt{nat}_i + \beta_2 \texttt{mortinf}_i + \beta_3 \texttt{lpnb}_i + \epsilon_i,$$

donde los errores ϵ_i son v.a.i.i.d. $\mathrm{N}(0, \sigma^2)$, se ha utilizado el programa `R` con los resultados siguientes:

```
reg = lm(Y~nat+mortinf+lpnb)
summary(reg)
Call:
lm(formula = Y ~ nat + mortinf + lpnb)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 69.24045  2.90253 23.855 < 2e-16
nat         -0.17572  0.04244 -4.140   8e-05
mortinf     -0.14086  0.01370 -10.284 < 2e-16
lpnb         0.98901  0.29404  3.363 0.00115
---
Residual standard error: 2.788 on 87 degrees of freedom
Multiple R-Squared: 0.9303,  Adjusted R-squared: 0.9279
F-statistic: 386.9 on 3 and 87 DF, p-value: < 2.2e-16

anova(reg)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value Pr(>F)
nat        1 7602.7 7602.7 977.798 < 2.2e-16
mortinf    1 1334.2 1334.2 171.599 < 2.2e-16
lpnb       1   88.0   88.0  11.313 0.001146
Residuals 87 676.5    7.8
```

a) ¿De cuántos países consta la muestra utilizada?
b) ¿Cuál es la suma de cuadrados del modelo de regresión (MSS) que se utiliza para medir la variabilidad explicada por las tres variables regresoras?
c) ¿Cuánto vale la varianza muestral de la variable respuesta, $\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$?
d) Contrasta a nivel $\alpha = 0.05$ la hipótesis nula $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.
e) Determina cuál es la hipótesis nula y la alternativa correspondiente a cada uno de los tres estadísticos `F` que aparecen en la tabla de análisis de la varianza anterior.

Solución:

a) $n = 91$
b) MSS $= 7602.7 + 1334.2 + 88.0 = 9024.9$ (también se puede calcular a partir del coeficiente de determinación $R^2 = 0.9303$ y la RSS $= 676.5$)
c) TSS $=$ MSS $+$ RSS $= 9701.4$. También se puede calcular a partir del coeficiente de determinación y la RSS: TSS $=$ RSS$/(1 - R^2)$.
d) `p-value: <2.2e-16` $\Rightarrow$ Rechazamos $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ a cualquiera de los niveles habituales, en particular al nivel del 5 %.
e) Se van introduciendo de uno en uno los regresores (en el orden dado por la fórmula `Y nat + mortinf + lpnb`) y se van haciendo contrastes para ver si merece la pena introducir ese regresor, condicionado a la presencia en el modelo de los regresores anteriores y teniendo en cuenta la varianza residual del modelo con las tres variables explicativas.

Línea	Contraste
`nat`	$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ (sin regresores)
	$H_1 : \beta_1 \neq 0, \beta_2 = 0, \beta_3 = 0$ (sólo variable `nat`)
`mortinf`	$H_0 : \beta_2 = 0, \beta_3 = 0$ (sólo variable `nat`)
	$H_1 : \beta_2 \neq 0, \beta_3 = 0$ (variables `nat` y `mortinf` en el modelo)
`lpnb`	$H_0 : \beta_3 = 0$ (variables `nat` y `mortinf` en el modelo)
	$H_1 : \beta_3 \neq 0$ (los tres regresores en el modelo)