Se sospecha que ciertos riesgos ambientales y laborales pueden alterar la proporción de nacimientos de varones. En particular, hay controversia respecto al efecto que la exposición a radiación ionizante tiene sobre el *sex ratio* o cociente del número de nacimientos masculinos respecto a aquél de nacimientos femeninos. Scherb *et al.* (2013)[1] estudiaron el efecto sobre el *sex ratio* del accidente en la Central Nuclear de Chernóbil (Ucrania) en abril de 1986. Concretamente, desde 1959 hasta 1986 (ambos años incluidos) nacieron en Rusia 31 706 752 varones de un total de 61 890 823 nacimientos. Desde 1987 hasta el 2010 (ambos años incluidos) en Rusia nacieron 38 360 531 bebés de los que 19 727 605 fueron varones.

a) A nivel $\alpha = 0.05$, ¿hay suficiente evidencia estadística de que el accidente de Chernóbil redujo la proporción de nacimientos de niñas (sobre el total de nacimientos) en Rusia?.

b) Explica el siguiente código de R e interpreta la correspondiente salida:

```
prop.test(c(31706752,19727605),c(61890823,38360531),alternative="less",
           correct=FALSE)

        2-sample test for equality of proportions without continuity
   correction

data:  c(31706752, 19727605) out of c(61890823, 38360531)
X-squared = 366.74, df = 1, p-value < 2.2e-16
alternative hypothesis: less
95 percent confidence interval:
 -1.000000000 -0.001798013
sample estimates:
   prop 1    prop 2
0.5123013 0.5142683
```

c) A nivel de significación del 5 %, ¿hay evidencia a favor de la hipótesis de que la proporción de nacimientos de varones sobre el total de nacimientos después del accidente nuclear es diferente de la proporción análoga a nivel mundial, que es de 0.517?

Solución: Las v.a.'s consideradas en este problema son

$$X = \begin{cases} 1 & \text{si un nacimiento antes del accidente es de varón;} \\ 0 & \text{si el nacimiento es de mujer.} \end{cases}$$

$$Y = \begin{cases} 1 & \text{si un nacimiento después del accidente es de varón;} \\ 0 & \text{si el nacimiento es de mujer.} \end{cases}$$

La variable X sigue una distribución de probabilidad Bernoulli(p_1) y la Y sigue una Bernoulli(p_2). Los e.m.v. de los parámetros p_1 y p_2 son respectivamente $\hat{p}_1 = \bar{x} = 0.5123013$ y $p_2 = \bar{y} = 0.5142683$. Denotamos por n_1 y n_2 los tamaños muestrales de X e Y respectivamente. A lo largo de todo el ejercicio emplearemos la notación de este apartado.

a) A nivel $\alpha = 0.05$, nos piden realizar el contraste

$$\begin{aligned} H_0 :&\ \ p_1 \geq p_2 \\ H_1 :&\ \ p_1 < p_2, \end{aligned}$$

cuya región de rechazo es

$$R = \left\{ \bar{x} - \bar{y} < z_{0.95} \sqrt{\bar{p}(1-\bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right\} = \{ z < z_{0.95} \},$$

[1]Fuente: Scherb, H. Kusmierz, R. y Voigt, K. (2013). Increased sex ratio in Russia and Cuba after Chernobyl: a radiological hypothesis. *Environmental Health*, 12:63.

donde el estadístico del contraste es

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{y} \qquad z_{0.95} = -z_{0.05} \simeq -1.645.$$

Calculamos

$$\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 1.05 \cdot 10^{-8} \qquad \text{y} \qquad z = -19.151.$$

Por tanto, rechazamos H_0 a nivel $\alpha = 0.05$.

b) `prop.test` es la función para comparar proporciones. Concretamente, el código de R que aparece en el enunciado realiza el mismo contraste de hipótesis que en el apartado (a). El resultado del código es el estadístico del contraste `X-squared = 366.74`, que es simplemente el estadístico del contraste z calculado en (a) y elevado al cuadrado, es decir, `X-squared` $= z^2$. El p-valor del contraste es menor que $2.2 \cdot 10^{-6}$; por tanto, es mucho menor que cualquier de los niveles de significación α habituales y lo razonable es rechazar la hipótesis nula. Al final del resultado aparecen las medias muestrales de X e Y, $\hat{p}_1 = \bar{x} = 0.5123013$ y $\hat{p}_2 = \bar{y} = 0.5142683$ respectivamente.

c) Un intervalo de confianza al 95 % para la proporción de nacimientos de varones sobre el total de nacimientos después del accidente nuclear es

$$\mathrm{IC}_{95\,\%}(p_2) = \left(\bar{y} \mp z_{0.025}\sqrt{\frac{\bar{y}(1-\bar{y})}{n_2}}\right) = (0.5142683 \mp 0.0001582),$$

donde hemos usado que $z_{0.025} = 1.96$.

A nivel de significación del 5 %, queremos contrastar $H_0 : p_2 = 0.517$ frente a $H_1 : p_2 \neq 0.517$. Como $0.517 \notin \mathrm{IC}_{95\,\%}(p_2)$, rechazamos H_0 a nivel $\alpha = 0.05$.

También podemos hacer el contraste con R:

```
prop.test(19727605,38360531,correct=F)

        1-sample proportions test without continuity correction

data:  19727605 out of 38360531, null probability 0.5
X-squared = 31238, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5141101 0.5144265
sample estimates:
        p
0.5142683
```