6 Inferencia Estadística III: Contrastes de hipótesis.

Problemas

1. Una de las condiciones para que el agua de abasto pueda considerarse potable es que la concentración media de sulfatos no exceda las 250 ppm. Se ha obtenido una muestra de 10 observaciones de esta variable en una tubería de la red de abasto, con los siguientes valores: 232.13, 282.24, 261.98, 291.41, 283.44, 216.68, 211.35, 258.78, 303.89 y 272.70 ppm. ¿Puede aceptarse, con una significación del 5 % que el agua que circula por esta tubería es potable? ¿Cuál es la potencia de este contraste si la concentración media real es de 280 ppm?

Solución:
El agua es potable si la concentración media de sulfatos μ es inferior a 250 ppm. Por tanto, planteamos el contraste:

$$\begin{cases} H_0: & \mu \leq 250 \\ H_1: & \mu > 250 \end{cases}$$

La regla de decisión es *rechazar* H_0 *si* $t_{exp} > t_{n-1,\alpha}$, siendo:

$$t_{exp} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Utilizamos R para calcular media y desviación típica de la muestra:

```
> sulfatos = c(232.13, 282.24, 261.98, 291.41, 283.44,
     216.68, 211.35, 258.78, 303.89, 272.7)
> mean(sulfatos)

[1] 261.46

> sd(sulfatos)

[1] 31.837
```

El valor del estadístico de contraste es entonces:

```
> (texp = (mean(sulfatos) - 250)/(sd(sulfatos)/sqrt(10)))

[1] 1.1383
```

Buscamos en la tabla de la t de Student el valor $t_{9,0,025} = 2{,}262$. Como $t_{exp} \leq 2{,}262$ concluimos que, aunque la concentración de sulfatos registrada es ligeramente mayor que 250, ello puede deberse al azar y no hay evidencias suficientes para rechazar H_0 y puede por tanto aceptarse que el agua es potable. En R podíamos haber realizado directamente el contraste mediante:

```
> t.test(sulfatos, mu = 250, alternative = "greater")

        One Sample t-test

data:  sulfatos
t = 1.1383, df = 9, p-value = 0.1422
alternative hypothesis: true mean is greater than 250
95 percent confidence interval:
 243 Inf
sample estimates:
mean of x
   261.46
```

Como vemos, el p-valor es mayor que el nivel de significación $\alpha = 0{,}05$, por lo que puede aceptarse la hipótesis nula.

La potencia es la probabilidad de rechazar una hipótesis nula que es falsa. Dado que el problema pide calcular la potencia cuando la concentración real es de 280 ppm, hemos de calcular:

$$
\begin{aligned}
P\left(\text{Rechazar } H_0 \middle|\, \mu = 280\right) &= P\left(t_{exp} > 2{,}262 \middle|\, \mu = 280\right) = P\left(\frac{\bar{X} - 250}{S/\sqrt{n}} > 2{,}262 \middle|\, \mu = 280\right) = \\
&= P\left(\frac{\bar{X} - 280 + 280 - 250}{S/\sqrt{n}} > 2{,}262 \middle|\, \mu = 280\right) = \\
&= P\left(\frac{\bar{X} - 280}{S/\sqrt{n}} + \frac{280 - 250}{S/\sqrt{n}} > 2{,}262 \middle|\, \mu = 280\right) = \\
&= P\left(t_9 > 2{,}262 - \frac{30}{31{,}837/\sqrt{10}}\right) = \\
&= P\left(t_9 > -0{,}71778\right) = P\left(t_9 \leq 0{,}71778\right) = \\
&= \texttt{pt(0.71778,9)} = 0{,}75445
\end{aligned}
$$

Por tanto, si la concentración real de sulfatos está en 280 ppm, una regla de decisión como la actual, con un tamaño de muestra $n = 10$ tiene una probabilidad 0.75 de declarar como no potable agua que no es potable por ener 280 ppm de sulfatos. Si esta probabilidad es aceptable, la regla de decisión es adecuada. Si se estima que esta probabilidad es pequeña, debería tomarse una muestra de tamaño mayor.

2. El número diario de piezas fabricadas por una máquina A en cinco días ha sido: 50, 48, 53,

60, 37; mientras que, en esos mismos días, una máquina B ha fabricado 40, 51, 62, 55 y 64 piezas. Suponiendo que la distribución de los datos es aproximadamente normal, ¿puede afirmarse que, en media, la máquina B fabrica diariamente más piezas que la A?. Determinar cuál debería ser el tamaño muestral n de ambas muestras si se desea que la potencia del contraste para detectar una diferencia de al menos diez piezas entre ambas máquinas sea del 90 %

Solución:
El contraste a realizar en este caso es de la forma:

$$\begin{cases} H_0: & \mu_A \geq \mu_B \\ H_1: & \mu_A < \mu_B \end{cases}$$

La regla de decisión es *rechazar* H_0 *si* $t_{exp} < -t_{n,\alpha}$, siendo en este caso:

$$t_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad n = REDONDEO\left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}\right]$$

Utilizamos R para realizar los cálculos:

```
> A = c(50, 48, 53, 60, 37)
> B = c(40, 51, 62, 55, 64)
> t.test(A, B, alternative = "less")

        Welch Two Sample t-test

data:  A and B
t = -0.8417, df = 7.856, p-value = 0.2124
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
   -Inf 5.8295
sample estimates:
mean of x mean of y
     49.6      54.4
```

Para utilizar la regla de decisión enunciada más arriba, buscamos en la tabla de la t de Student el valor $t_{8,0{,}05} = 1{,}86$; como $t_{exp} = -0{,}8417 > -t_{8,0{,}05}$, no se da la condición para el rechazo de H_0 por lo que concluimos que aunque en estas muestras, la máquina A ha fabricado, por término medio, menos piezas que la máquina B, *la evidencia no es suficiente para asegurar que, en general, B fabrica más piezas que A* (la diferencia observada es atribuíble al azar). De modo equivalente, el contraste podía haberse llevado a cabo simplemente comparando el p-valor con el nivel de significación α=0.05. Como el p-valor es mayor que α, se acepta H_0.

El tamaño de muestra necesario para realizar este contraste de tal forma que la potencia para detectar una diferencia de al menos $\Delta = 10$ piezas entre ambas máquinas sea del 90 % es:

$$n = \frac{(z_\alpha + z_\beta)^2 \left(s_1^2 + s_2^2\right)}{\Delta^2}$$

Calculamos los valores necesarios:

```
> var(A)

[1] 70.3

> var(B)

[1] 92.3

> (za = qnorm(0.95))

[1] 1.6449

> (zb = qnorm(0.9))

[1] 1.2816

> delta = 10
> n = ((za + zb)^2) * (var(A) + var(B))/delta^2
> n

[1] 13.925
```

Por tanto, se necesitaría una muestra de al menos 14 observaciones (producción de 14 días) en cada máquina.

3. Se realizan 10 determinaciones del porcentaje de riqueza en un polímero con dos instrumentos distintos. Las varianzas muestrales resultan ser 0.5419 y 0.6065. ¿Existe evidencia suficiente al 5 % de significación para asegurar que el segundo instrumento presenta más variabilidad en las medidas que el primero?

Solución:
Hemos de realizar el contraste:

$$\begin{cases} H_0: & \sigma_1^2 \geq \sigma_2^2 \\ H_1: & \sigma_1^2 < \sigma_2^2 \end{cases}$$

Los datos muestran que la varianza muestral con el segundo instrumento es mayor que con el primero. Para determinar si puede generalizarse este resultado hemos de comprobar si $F_{exp} < F_{n_1-1, n_2-1, 1-\alpha}$, siendo $F_{exp} = \frac{S_1^2}{S_2^2} = \frac{0{,}5419}{0{,}6065} = 0{,}89349$. En la tabla de la F de Fisher encontramos que $F_{9,9,0,95} = \frac{1}{F_{9,9,0,05}} = \frac{1}{3,18} = 0,31447$. Como $F_{exp} > F_{9,9,0,05}$, concluimos que no se da la condición para el rechazo de H_0. Así pues, no hay evidencia suficiente de que el segundo instrumento presenta más variabilidad que el primero.

NOTA: hemos supuesto aquí que las dos muestras en las que se basa este análisis son independientes; si fuesen emparejadas debería utilizarse el test de Pitman-Morgan:

$$t_{exp} = \frac{(F-1)\sqrt{n-2}}{2\sqrt{F\,(1-r^2)}}$$

donde F es el cociente entre la mayor y menor de las dos varianzas, n es el número de parejas de datos y r es la correlación. Obviamente, con los datos aportados en este problema no puede llevarse a cabo el contraste (falta la correlación). El valor de la t_{exp} debe compararse, para tomar la decisión, con una t con $n-2$ grados de libertad. En R este contraste se encuentra en la librería `PairedData`.

4. En una encuesta sobre el hábito de fumar se ha constatado que, de 900 personas entrevistadas en la ciudad A, 554 son fumadoras, mientras que en otra ciudad B, entre 600 entrevistadas se han encontrado 410 fumadoras. ¿Es significativamente distinta la proporción de fumadores en ambas ciudades?

Solución:

Sean π_A la proporción de fumadores en A y π_B la proporción en B. Hemos de llevar a cabo el contraste:

$$\begin{cases} H_0: & \pi_A = \pi_B \\ H_1: & \pi_A \neq \pi_B \end{cases}$$

Para ello disponemos de dos muestras independientes de tamaños respectivos $n_A = 900$ y $n_B = 600$. Las proporciones estimadas de fumadores en cada ciudad son, respectivamente, $p_A = \frac{554}{900} = 0,61556$ y $p_B = \frac{410}{600} = 0,68333$. Dado que los tamaños muestrales son suficientemente grandes y siendo:

$$p^* = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{554 + 410}{900 + 600} = \frac{964}{1500} = 0,64267, \quad q^* = 1 - p^* = 0,35733$$

las proporciones globales de fumadores y no fumadores, respectivamente, se tiene que $n_A p^*, n_A q^*, n_B p^*, n_B q^*$. Por tanto podemos aplicar el contraste basado en el estadístico:

$$z_{exp} = \frac{p_A - p_B}{\sqrt{p^* q^* \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = -2,6835$$

La regla de decisión consiste en rechazar H_0 si $|z_{exp}| > z_{\alpha/2}$. Para $\alpha = 0,05$ es $z_{0,025} = 1,96$, y como $|-2,6835| > 1,96$ concluimos que existe evidencia suficiente de que la proporción de fumadores difiere de manera significativa entre ambas ciudades. Podemos acompañar esta información por el intervalo de confianza:

$$\ln\left(\frac{\pi_A}{\pi_B}\right) \in \left[\ln\left(\frac{p_A}{p_B}\right) \pm z_{\alpha/2}\sqrt{\frac{(1-p_A)}{n_A p_A} + \frac{(1-p_B)}{n_B p_B}}\right] = [-0{,}17951, -0{,}029405]$$

o lo que es lo mismo:

$$\frac{\pi_A}{\pi_B} \in \left[e^{-0{,}17951}, e^{-0{,}029405}\right] = [0{,}83568, 0{,}97102]$$

lo que significa que la proporción de fumadores en A es entre un 83.568 % y un 97.102 % de la proporción de fumadores en B.

El contraste se resuelve fácilmente en R mediante:

```
> prop.test(c(554, 410), c(900, 600))

        2-sample test for equality of proportions with continuity
        correction

data:  c(554, 410) out of c(900, 600)
X-squared = 6.9093, df = 1, p-value = 0.008575
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.118110 -0.017445
sample estimates:
 prop 1  prop 2
0.61556 0.68333
```

R utiliza un método diferente del señalado aquí, si bien se llega a la misma conclusión (la diferencia entre proporciones es significativa) ya que el p-valor (0.0086) es menor que $\alpha = 0{,}05$. Asimismo, R muestra un intervalo para la diferencia de proporciones $\pi_A - \pi_B$ que es negativo, indicando con ello que $\pi_A < \pi_B$.

5. Se desea determinar si el diámetro medio de los granos de arena es el mismo en dos distintas franjas de una playa; para ello se dispone de los siguientes datos (en centésimas de mm.):

ZONA A	22	24	25	22	28	26	25	24	28	26	23	23
ZONA B	28	27	27	29	26	29	27	30	28	27	26	29

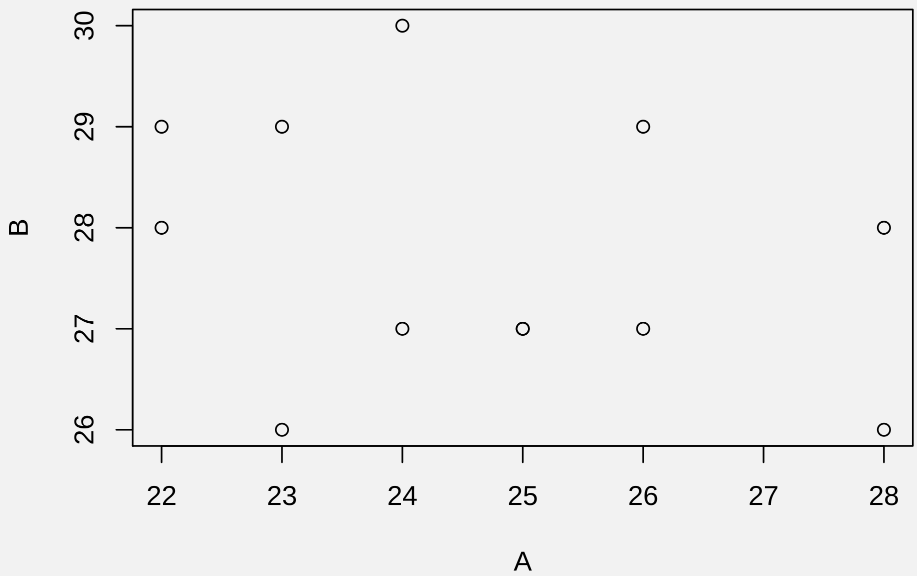
Decidir, con un nivel de significación del 5 % si el diámetro medio es el mismo en las dos zonas.

Solución:

El problema no nos informa de si los datos están emparejados o no (puede ser que los datos de

A y B se hayan cogido, en cada caso, en la misma franja perpendicular a la costa). Un simple gráfico nos puede ayudar a decidir si hay o no emparejamiento:

```
> A = c(22, 24, 25, 22, 28, 26, 25, 24, 28, 26, 23, 23)
> B = c(28, 27, 27, 29, 26, 29, 27, 30, 28, 27, 26, 29)
> plot(A, B)
```



Este gráfico muestra claramente que no hay asociación entre A y B, por lo que podemos considerar ambas muestras como independientes. En caso de duda podríamos hacer previamente un contraste sobre la correlación entre ambas variables:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

que se resuelve calculando:

$$t_{\text{exp}} = \sqrt{\frac{(n-2)r^2}{1-r^2}}$$

y si $t_{exp} \leq t_{n-2,\alpha/2}$ se acepta H_0. En este caso, $n = 12$ y $t_{10,0{,}025} = 2{,}228$. Calculamos r y t_{exp}:

```
> (r = cor(A, B))
[1] -0.3084
> (texp = sqrt(10 * r^2/(1 - r^2)))
[1] 1.0252
```

Como $t_{exp} \leq 2{,}228$ concluimos (como ya era evidente en el gráfico) que no existe asociación entre A y B. En R este contraste puede realizarse fácilmente mediante:

```
> cor.test(A, B)

        Pearson's product-moment correlation

data:  A and B
t = -1.0252, df = 10, p-value = 0.3294
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.74962  0.32260
sample estimates:
    cor
-0.3084
```

El p-valor mayor que $\alpha = 0{,}05$ indica que la correlación no es significativa (esto es, no difiere significativamente de cero o, dicho en otras palabras, a los efectos prácticos la correlación es nula, no hay asociación entre los valores en A y B).

Dicho todo lo anterior, para decidir si el diámetro medio de los granos de arena es el mismo en las dos zonas, resolvemos el contraste:

$$\begin{cases} H_0: & \mu_A = \mu_B \\ H_1: & \mu_A \neq \mu_B \end{cases}$$

considerando muestras independientes. Para ello calculamos el estadístico:

$$t_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

y la regla de decisión consiste en rechazar H_0 si $|t_{exp}| > t_{n,\alpha/2}$, siendo:

$$n = REDONDEO\left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}\right]$$

Tenemos:

```
> mean(A)

[1] 24.667

> mean(B)

[1] 27.75

> sd(A)

[1] 2.0597
```

```
> sd(B)

[1] 1.2881
```

Sustituyendo estos valores en las fórmulas anteriores tenemos que $t_{exp} = -4{,}3967$ y $n = 18$. En la tabla de la t de Student se obtiene $t_{18,0{,}025} = 2{,}101$. Como se cumple que $|t_{exp}| > t_{n,\alpha/2}$, concluimos que existe evidencia suficiente para asegurar que el diámetro medio de los granos de arena difiere significativamente entre las dos zonas de la playa.

El contraste se puede llevar a cabo de manera muy simple en R mediante:

```
> t.test(A, B)

        Welch Two Sample t-test


data:  A and B
t = -4.3967, df = 18.462, p-value = 0.0003295
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.5540 -1.6126
sample estimates:
mean of x mean of y
   24.667    27.750
```

El p-valor del test (0.00032955), menor que $\alpha = 0{,}05$ indica que, efectivamente, la diferencia entre los diámetros medios de los granos de arena en ambas zonas de la playa es significativa. La salida de R muestra además el intervalo de confianza para la magnitud de dicha diferencia, situada con un 95 % de confianza entre -4.554 y -1.6126 mm.

6. Se estudia la supervivencia de ratones de laboratorio a cierta enfermedad en función del tratamiento administrado; los siguientes datos muestran la supervivencia en días desde que se contrae la enfermedad:

TRATAMIENTO 1	29	42	38	40	43	40	30	42		
TRATAMIENTO 2	30	35	39	28	31	31	29	35	29	33

Hasta ahora el tratamiento aplicado preferentemente viene siendo el 2. Decidir si estos datos contienen evidencia suficiente de que el tratamiento 1 ofrece mayor tiempo medio de supervivencia que el 2.

Solución:

En este caso es obvio que se trata de dos muestras independientes (los ratones del primer grupo son distintos, y además son menos, que los del segundo). El contraste a realizar es:

$$\begin{cases} H_0: & \mu_1 \leq \mu_2 \\ H_1: & \mu_1 > \mu_2 \end{cases}$$

que debe resolverse mediante el estadístico:

$$t_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

siendo la regla de decisión rechazar H_0 si $t_{exp} > t_{n,\alpha}$, siendo:

$$n = REDONDEO\left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}\right]$$

Tenemos:

```
> trat1 = c(29, 42, 38, 40, 43, 40, 30, 42)
> trat2 = c(30, 35, 39, 28, 31, 31, 29, 35, 29, 33)
> mean(trat1)

[1] 38

> mean(trat2)

[1] 32

> sd(trat1)

[1] 5.4772

> sd(trat2)

[1] 3.4641
```

Sustituyendo estos valores en las fórmulas anteriores tenemos que $t_{exp} = 2{,}6968$ y $n = 11$. En la tabla de la t de Student se obtiene $t_{11,0{,}05} = 1{,}796$. Como se cumple que $t_{exp} > t_{n,\alpha}$, concluimos que existe evidencia suficiente para asegurar que el tratamiento 1 produce, en media, mayor tiempo de supervivencia que el tratamiento 2.

El contraste se puede llevar a cabo duna e manera muy simple en R mediante:

```
> t.test(trat1, trat2, alternative = "greater")

        Welch Two Sample t-test

data:  trat1 and trat2
t = 2.6968, df = 11.297, p-value = 0.01017
```

```
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.014   Inf
sample estimates:
mean of x mean of y
       38        32
```

El p-valor del test (0.010173), menor que $\alpha = 0{,}05$ indica que, efectivamente, la diferencia entre los tiempos medios de supervivencia bajo ambos tratamientos es significativa. El intervalo de confianza, entre 2.014 e ∞ indica además que, con un 95 % de confianza, el primer tratamiento produce una supervivencia media de al menos 2.014 días más que el segundo.

7. Se desea probar la precisión de una balanza diseñada para pesar a bordo de un barco. Para ello se eligen al azar diez objetos y se pesan con esa balanza en el barco, y posteriormente en tierra. Los pesos medidos para cada objeto figuran en la siguiente tabla (en kg.):

BARCO	0.14	0.20	0.07	0.18	0.38	0.10	0.04	0.27	0.27	0.21
TIERRA	0.16	0.19	0.09	0.20	0.34	0.10	0.08	0.29	0.28	0.17

¿Puede aceptarse que, en media, el peso medido en el barco es el mismo que en tierra?. ¿Cuál es la potencia del contraste para detectar una diferencia media mínima de 0.1 kg. entre las dos medidas? Si se deseara que esta potencia fuese del 90 % ¿cuál debería ser el tamaño de las muestras?

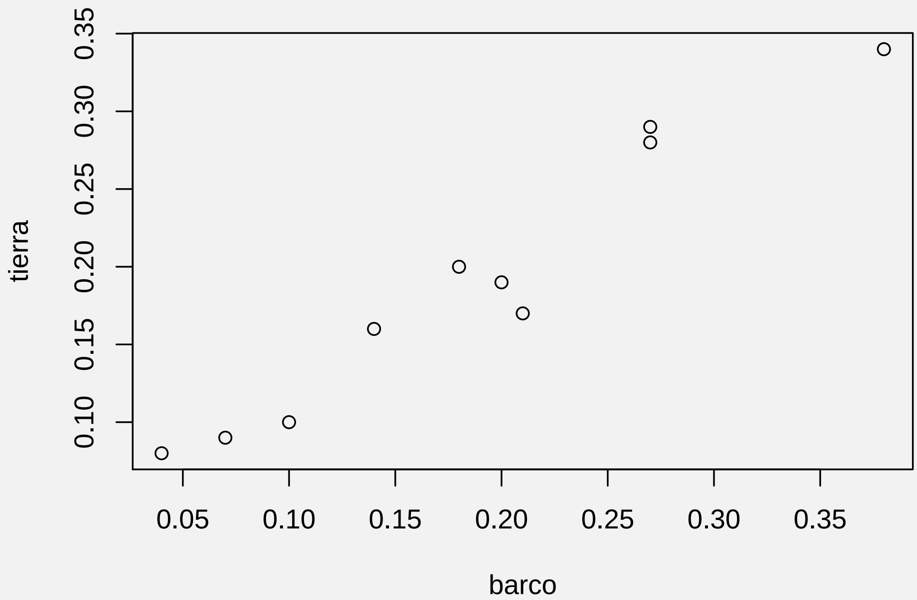
Solución:

El contraste a resolver es:

$$\begin{cases} H_0: & \mu_{BARCO} = \mu_{TIERRA} \\ H_1: & \mu_{BARCO} \neq \mu_{TIERRA} \end{cases}$$

Dado que es el mismo objeto el que se pesa en el barco y en tierra cabe esperar asociación entre ambos valores. El siguiente gráfico permite comprobar que tal asociación efectivamente existe:

```
> barco = c(0.14, 0.2, 0.07, 0.18, 0.38, 0.1, 0.04, 0.27,
     0.27, 0.21)
> tierra = c(0.16, 0.19, 0.09, 0.2, 0.34, 0.1, 0.08, 0.29,
     0.28, 0.17)
> plot(barco, tierra)
```



Por tanto consideramos las muestras como emparejadas y el contraste debe resolverse utilizando el estadístico:

$$t_{exp} = \frac{\overline{X}_1 - \overline{X}_2}{S_D/\sqrt{n}}$$

siendo $S_D = \sqrt{S_1^2 + S_2^2 - 2rS_1S_2}$ y n el número de parejas de datos; la regla de decisión consiste en rechazar H_0 si $|t_{exp}| > t_{n-1,\alpha/2}$. Podemos utilizar R para realizar los cálculos:

```
> (x1 = mean(barco))

[1] 0.186

> (x2 = mean(tierra))

[1] 0.19

> (s1 = sd(barco))

[1] 0.10373

> (s2 = sd(tierra))

[1] 0.089567

> (r = cor(barco, tierra))

[1] 0.97229
```

Como vemos, la diferencia en los valores medios entre el peso en tierra y el peso en el barco es muy pequeña. No obstante para determinar si es o no significativa, seguimos adelante con el contraste. Con los datos anteriores:

```
> (sD = sqrt(s1^2 + s2^2 - 2 * r * s1 * s2))

[1] 0.02675

> (texp = (x1 - x2)/(sD/sqrt(10)))

[1] -0.47287
```

El valor $t_{n-1,\alpha/2} = t_{9,0{,}025}$ podemos buscarlo en la tabla de la t de Student o también utilizar R :

$$t_{9,0{,}025} = \texttt{qt(0.975,9)} = 2{,}2622$$

Como resulta que $|t_{exp}| = 0{,}47287 \leq t_{9,0{,}025}$, puede aceptarse H_0. Concluimos por tanto que, efectivamente, la diferencia entre pesos medios no es significativa y puede atribuirse simplemente al azar (error experimental). En R este contraste puede realizarse directamente mediante:

```
> t.test(tierra, barco, paired = TRUE)

        Paired t-test

data:  tierra and barco
t = 0.4729, df = 9, p-value = 0.6476
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.015136  0.023136
sample estimates:
mean of the differences
                  0.004
```

El p-valor (0.64757) es mucho mayor que $\alpha = 0{,}05$, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula, por lo que no existe diferencia significativa entre el peso medio obtenido en el barco y el obtenido en tierra.

Para responder a la cuestión de cuál es la potencia del contraste para detectar una diferencia media mínima de 0.1 kg. entre las dos medidas, recordemos que la potencia es la probabilidad de rechazar H_0 cuando es falsa; en este caso concreto, cuando es falsa porque existe una diferencia de al menos 0.1 kg:

$$\begin{aligned} Potencia\,(0{,}1\,kg) &= P\,(Rechazar\,H_0\,|\mu_D = 0{,}1\,) = P\,(|t_{exp}| \geq t_{9,0{,}025}\,|\mu_D = 0{,}1\,) = \\ &= 1 - P\,(|t_{exp}| \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\,) = 1 - P\left(\left|\frac{\overline{X}_1 - \overline{X}_2}{S_D/\sqrt{n}}\right| \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\right) \end{aligned}$$

Ahora bien, cuando $\mu_D = 0{,}1$, se tiene que $\frac{\overline{X}_1 - \overline{X}_2 - \mu_D}{S_D/\sqrt{n}} \approx t_9$. Entonces:

$$\begin{aligned}
1- \quad & P\left(\left|\frac{\overline{X}_1 - \overline{X}_2}{S_D/\sqrt{n}}\right| \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\right) = \\
= \quad & 1 - P\left(-t_{9,0{,}025} \leq \frac{\overline{X}_1 - \overline{X}_2}{S_D/\sqrt{n}} \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\right) = \\
= \quad & 1 - P\left(-t_{9,0{,}025} \leq \frac{\overline{X}_1 - \overline{X}_2 - 0{,}1 + 0{,}1}{S_D/\sqrt{n}} \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\right) = \\
= \quad & 1 - P\left(-t_{9,0{,}025} \leq \frac{\overline{X}_1 - \overline{X}_2 - 0{,}1}{S_D/\sqrt{n}} + \frac{0{,}1}{S_D/\sqrt{n}} \leq t_{9,0{,}025}\,|\mu_D = 0{,}1\right) = \\
= \quad & 1 - P\left(-t_{9,0{,}025} - \frac{0{,}1}{S_D/\sqrt{n}} \leq \frac{\overline{X}_1 - \overline{X}_2 - 0{,}1}{S_D/\sqrt{n}} \leq t_{9,0{,}025} - \frac{0{,}1}{S_D/\sqrt{n}}\,|\mu_D = 0{,}1\right) = \\
= \quad & 1 - P\left(-2{,}2622 - \frac{0{,}1}{0{,}02675/\sqrt{10}} \leq t_9 \leq 2{,}2622 - \frac{0{,}1}{0{,}02675/\sqrt{10}}\right) = \\
= \quad & 1 - P\left(-14{,}084 \leq t_9 \leq -9{,}5595\right) = \\
= \quad & 1 - \left(P\left(t_9 \leq -9{,}5595\right) - P\left(t_9 \leq -14{,}084\right)\right) = \\
= \quad & 1 - (\texttt{pt(-9.5595,9)} - \texttt{pt(-14.084,9)}) = \\
= \quad & 1
\end{aligned}$$

Por tanto la potencia es 1, es decir, con seguridad se rechazaría la hipótesis nula (que ambos pesos son iguales) si la diferencia entre el peso en tierra y en el barco superase los 0.1 kg. Para calcular el tamaño de la muestra si se desea que esta potencia sea $1 - \beta = 0{,}9$, con $\alpha = 0{,}05$ empleamos la fórmula:

$$n = \frac{(z_{\alpha/2} + z_\beta)^2 s_D^2}{\Delta^2} = \frac{(1{,}96 + 1{,}2816)^2\, 0{,}02675^2}{0{,}1^2} = 0{,}75186 \cong 1$$

En otras palabras, la variabilidad en la diferencia de pesos es tan pequeña en comparación con la diferencia que se pretende detectar, que con una simple muestra de tamaño 1 se puede tomar la decisión con los requisitos de significación y potencia indicados. Si observamos las diferencias entre pesos:

```
> barco - tierra
```

```
 [1] -0.02  0.01 -0.02 -0.02  0.04  0.00 -0.04 -0.02 -0.01  0.04
```

vemos que estas diferencias son del orden de centésimas, con lo cual cualquier muestra de tamaño 1 sería suficiente para detectar una diferencia de una décima. Si la diferencia a detectar fuese de una centésima (0.01), el tamaño de muestra necesario sería notablemente mayor:

$$n = \frac{(1{,}96 + 1{,}2816)^2\, 0{,}02675^2}{0{,}01^2} = 75{,}186 \cong 76$$

8. Dos industrias químicas vierten sus residuos directamente al mar. Se han tomado datos sobre

el contenido en fenoles de estos residuos, eligiéndose para ello doce muestras de residuos procedentes de la primera industria y diez procedentes de la segunda. Los resultados obtenidos en los análisis se muestran a continuación (en ‰ de fenoles por cada 100 gr de peso seco de los residuos):

A	28	32	45	23	34	21	38	19	41	23	18	24
B	35	23	20	17	14	42	18	20	22	19		

Decidir, con un nivel de significación del 5 % si existen diferencias significativas en las cantidades de fenoles producidos por ambas industrias. De acuerdo con la normativa legal, el valor medio máximo de fenoles que puede ser vertido al mar sin tratamiento es del 30‰. ¿Cumplen estas empresas con la normativa?. Téngase en cuenta para esta última pregunta, que se desea una probabilidad muy baja (0,01) de aceptar que se cumple la normativa cuando no se cumple.

Solución:

La primera cuestión requiere llevar a cabo el contraste

$$\begin{cases} H_0: & \mu_A = \mu_B \\ H_1: & \mu_A \neq \mu_B \end{cases}$$

siendo ambas muestras independientes. Lo hacemos directamente en R :

```
> A = c(28, 32, 45, 23, 34, 21, 38, 19, 41, 23, 18, 24)
> B = c(35, 23, 20, 17, 14, 42, 18, 20, 22, 19)
> t.test(A, B)

        Welch Two Sample t-test

data:  A and B
t = 1.5384, df = 19.532, p-value = 0.14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0885 13.7552
sample estimates:
mean of x mean of y
   28.833    23.000
```

El p-valor del contraste (0.14) indica que la diferencia observada (algo más de 5 unidades) no es significativa. Para determinar si las empresas cumplen con la normativa, dado que se desea una probabilidad muy baja (0,01) de aceptar que se cumple la normativa cuando no se cumple, hemos de realizar un contraste de la forma:

$$\begin{cases} H_0: & \mu \geq 30 \\ H_1: & \mu < 30 \end{cases}$$

con $\alpha = 0{,}01$. Como en casos anteriores, habría que calcular el estadístico:

$$t_{exp} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

con $\mu_0 = 30$ y rechazar la hipótesis nula si $t_{exp} < -t_{n-1,\alpha}$. En este caso vamos a utilizar directamente R y emplear la regla del p-valor:

```
> t.test(A, mu = 30, alternative = "less")

        One Sample t-test

data:  A
t = -0.4474, df = 11, p-value = 0.3316
alternative hypothesis: true mean is less than 30
95 percent confidence interval:
   -Inf 33.517
sample estimates:
mean of x
   28.833

> t.test(B, mu = 30, alternative = "less")

        One Sample t-test

data:  B
t = -2.5429, df = 9, p-value = 0.01578
alternative hypothesis: true mean is less than 30
95 percent confidence interval:
   -Inf 28.046
sample estimates:
mean of x
       23
```

En el caso de la empresa B hay una evidencia fuerte de que está cumpliendo la normativa (tasa de emisión de fenoles significativamente por debajo del 30‰), ya que el p-valor (0.01578) es menor que 0.05 y por tanto lleva al rechazo de H_0. En cambio, para la empresa A, aunque en la muestra la tasa de emisión de fenoles está por debajo del 30‰ la evidencia no es suficiente para afirmar que cumple la normativa.

Posiblemente desde el punto de vista de la empresa se puede decir que esta manera de hacer las cosas es demasiado exigente ya que, si bien es poco probable decidir que se cumple la normativa cuando no se cumple, a cambio es posible que sea muy probable decidir que no se cumple cuando de hecho sí que se está cumpliendo. Desde este punto de vista, si la empresa A estuviera emitiendo un 29‰ (que estaría dentro de lo permitido), la probabilidad de que el método aplicado indicase que no cumple sería (en estos cálculos necesitamos la desviación típica de las emisiones de A, `sd(A)`=9.0336):

$$
\begin{aligned}
P\left(aceptar\ H_0 \middle|\, \mu = 29\right) &= P\left(\frac{\bar{X}-30}{S/\sqrt{n}} \geq -t_{11,0,05} \middle|\, \mu = 29\right) = \\
&= P\left(\frac{\bar{X}-29+29-30}{S/\sqrt{n}} \geq -t_{11,0,05} \middle|\, \mu = 29\right) = \\
&= P\left(\frac{\bar{X}-29}{S/\sqrt{n}} - \frac{1}{S/\sqrt{n}} \geq -t_{11,0,05} \middle|\, \mu = 29\right) = \\
&= P\left(t_{11} - \frac{1}{9{,}0336/\sqrt{12}} \geq -1{,}7959\right) = \\
&= P\left(t_{11} \geq -1{,}4124\right) = P\left(t_{11} \leq 1{,}4124\right) = \\
&= \texttt{pt(1.4124,11)} = 0{,}90726
\end{aligned}
$$

Así pues, en este caso la empresa cumple, pero hay una probabilidad del 90 % de que el test diga que no lo hace. Si se desea que esta probabilidad sea menor (por ejemplo, que sea sólo del 5 % cuando $\mu = 29$) el tamaño de la muestra debería aumentar hasta:

$$
n = \frac{(z_\alpha + z_\beta)^2 s^2}{\Delta^2} = \frac{\left(z_{0,01} + z_{0,05}\right)^2 9{,}0336^2}{1} = \frac{\left(2{,}3263 + 1{,}6449\right)^2 9{,}0336^2}{1^2} \cong 1287
$$

lo cual probablemente haría impracticable el muestreo.

9. Con objeto de determinar cuál de dos diferentes técnicas de cultivo de peces produce mayor rendimiento, se ha medido la producción, en toneladas, durante 12 periodos para la técnica A y durante 10 periodos para la B, con los siguientes resultados:

A	8.93	9.54	10.32	6.99	8.56	8.67	9.72	7.76	8.95	9.32	8.59	9.78
B	5.69	7.56	6.88	10.26	9.57	7.88	8.95	9.35	6.58	7.32		

¿Existe evidencia de que el rendimiento medio obtenido con la técnica A es superior al obtenido con la B? Responder a esta cuestión: a) suponiendo que los datos siguen una distribución normal. b) suponiendo que los datos no siguen una distribución normal.

Solución:
En el caso de que los datos sigan distribución normal debe emplearse el test de la t de Student; en caso contrario, el método adecuado es el test de Wilcoxon. Aplicamos R directamente a los valores observados:

```
> A = c(8.93, 9.54, 10.32, 6.99, 8.56, 8.67, 9.72, 7.76,
     8.95, 9.32, 8.59, 9.78)
> B = c(5.69, 7.56, 6.88, 10.26, 9.57, 7.88, 8.95, 9.35,
     6.58, 7.32)
> t.test(A, B, alternative = "greater")
```

```
        Welch Two Sample t-test

data:  A and B
t = 1.722, df = 14.505, p-value = 0.05316
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.018769       Inf
sample estimates:
mean of x mean of y 
   8.9275    8.0040 
```

```
> wilcox.test(A, B, alternative = "greater")
```

```
        Wilcoxon rank sum test with continuity correction

data:  A and B
W = 81.5, p-value = 0.08301
alternative hypothesis: true location shift is greater than 0
```

Como vemos, el rendimiento medio obtenido en la muestra A (8.9275) es algo mayor que el obtenido con la B (8.0040). No obstante, tanto si se supone normalidad como si no, la evidencia es insuficiente para asegurar que en general A presenta mayor rendimiento que B (los p-valores son mayores que $\alpha = 0{,}05$ en ambos casos: p-valor 0.05316 en caso de asumir normalidad, 0.08301 en caso de no asumirla).

Si deseamos decidir si puede asumirse o no la normalidad de la distribución de los datos en ambas muestras, realizamos el contraste de Shapiro-Wilk:

$$\begin{cases} H_0: & X \approx N(\mu, \sigma) \\ H_1: & X \not\approx N(\mu, \sigma) \end{cases}$$

```
> shapiro.test(A)
```

```
        Shapiro-Wilk normality test

data:  A
W = 0.9584, p-value = 0.7611
```

```
> shapiro.test(B)
```

```
        Shapiro-Wilk normality test

data:  B
W = 0.9658, p-value = 0.849
```

Como vemos, en ambos casos los p-valores superan el nivel de significación $\alpha = 0{,}05$, por lo que puede aceptarse la normalidad de los datos de ambas muestras.

10. Un laboratorio químico desarrolla un nuevo insecticida para eliminar las plagas de pulgón en cultivos de cereales. El insecticida comercializado hasta ahora por el laboratorio tiene una eficacia del 60 % (esto es, elimina al 60 % de los pulgones sobre los que actúa, en un plazo inferior a 24 horas). Como prueba preliminar, el nuevo insecticida se ensaya sobre 14 pulgones, resultando nueve de ellos muertos antes de 24 horas ¿Es ésta evidencia suficiente de que el nuevo insecticida es más eficaz que el anterior? Posteriormente el insecticida se ensaya sobre 83 pulgones, muriendo 55 antes de las 24 horas. ¿Existe ahora evidencia de que el nuevo insecticida es más eficaz que el anterior?

Solución:
Para responder a la primera cuestión hemos de resolver el contraste:

$$\begin{cases} H_0: & \pi = 0{,}6 \\ H_1: & \pi > 0{,}6 \end{cases}$$

Como el tamaño de muestra es pequeño $(n = 14)$ utilizamos como estadístico de contraste el número de éxitos N_E (en este caso el número de pulgones muertos), siendo la regla de decisión *rechazar* H_0 *si* $N_E > b_{n,\alpha}(\pi_0)$, con:

$$b_{n,\alpha}(\pi_0) = \text{mín}\{k \mid P(X > k) \leq \alpha\} = \text{mín}\{k \mid P(X \leq k) \geq 1-\alpha\} = \texttt{qbinom(}1-\alpha\texttt{,n,}\pi_0\texttt{)}$$

El número de pulgones muertos en este experimento fue $N_E = 9$ de $n = 14$ (eficacia $\frac{9}{14} = 0{,}64286 \cong 64{,}286\,\%$). Utilizando R resulta $b_{14,0{,}05}(0{,}6) =$ `qbinom(0.95,14,0.6)` $= 11$. Como $N_E = 9 < b_{14,0{,}05}(0{,}6)$, no se da la condición para rechazar H_0, por lo que aunque la eficacia es este experimento ha sido algo mayor del 60 %, la evidencia es insuficiente para asegurar con un nivel de significación del 5 % que el nuevo insecticida es más eficaz que el anterior. Este contraste puede realizarse directamente en R mediante:

```
> binom.test(9, 14, 0.6, alternative = "greater")

        Exact binomial test


data:  9 and 14
number of successes = 9, number of trials = 14, p-value =
0.4859
alternative hypothesis: true probability of success is greater than 0.6
95 percent confidence interval:
 0.39041 1.00000
sample estimates:
probability of success
               0.64286
```

El p-valor de este contraste (0.4859) muestra claramente que no puede rechazarse la hipótesis nula, y por tanto que no existe evidencia suficiente para asegurar que el nuevo inseticida tiene una eficacia superior al 60 %. Observando el intervalo de confianza, con los datos disponibles, lo más que podemos llegar a asegurar con una confianza del 95 % es que el nuevo insecticida tiene una eficacia de al menos el 39.041 %.

Cuando posteriormente se repite el experimento sobre 83 pulgones, con 55 muertos, la nueva tasa de eficacia estimada es $\frac{55}{83} = 0{,}66265 \cong 66{,}265\,\%$ (que se supone más precisa que la anterior dado que se ha estimado con más datos). Repetimos el contraste anterior con estos nuevos datos y obtenemos:

```
> binom.test(55, 83, 0.6, alternative = "greater")

        Exact binomial test

data:  55 and 83
number of successes = 55, number of trials = 83, p-value =
0.1459
alternative hypothesis: true probability of success is greater than 0.6
95 percent confidence interval:
 0.56781 1.00000
sample estimates:
probability of success
               0.66265
```

Aunque la eficacia en la muestra vuelve a superar el 60 %, la evidencia sigue siendo insuficiente para asegurar con una significación del 5 % que la eficacia de este nuevo insecticida supera en general el valor del 60 %, ya que el p-valor (0.1459) sigue siendo mayor que α=0.05.

NOTA: este segundo apartado podía haberse resuelto “a mano” utilizando la aproximación $z_{exp} = \frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$, ya que $n \geq 30$, $N_E > 5$ y $n - N_E > 5$. En este caso:

$$z_{exp} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}66265 - 0{,}6}{\sqrt{\frac{0{,}6(1-0{,}6)}{83}}} = 1{,}1651$$

y como $z_{exp} < z_{0{,}05} = 1{,}96$ concluimos que no existen evidencias suficientes para rechazar H_0 (en otras palabras, el nuevo insecticida no es más eficaz que el anterior).

11. Los pulgones que atacan a la cebada constituyen una variedad diferente de la que ataca al trigo. El nuevo insecticida se ha desarrollado de tal forma que se espera que sea absolutamente genérico y que por tanto su eficacia sea la misma para ambas variedades de pulgón. Se ha aplicado el insecticida a 16 ejemplares de pulgón del trigo y a 19 de pulgón de la cebada. Al cabo de 24 horas han muerto 5 pulgones de la primera clase y 10 de la segunda. ¿Es ésta evidencia suficiente en contra de la hipótesis de que el insecticida es genérico?

Solución:

El contraste a resolver es de la forma:

$$\begin{cases} H_0: & \pi_{Trigo} = \pi_{Cebada} \\ H_1: & \pi_{Trigo} \neq \pi_{Cebada} \end{cases}$$

Nuestras estimaciones a partir de los datos aportados por las muestras son:

$$p_{Trigo} = \frac{5}{16} = 0{,}3125 \qquad p_{Cebada} = \frac{10}{19} = 0{,}52632$$

Obviamente los valores son distintos; para determinar si la evidencia de que el efecto es distinto sobre ambas variedades de pulgón es suficiente para poder generalizarla a las poblaciones respectivas debemos llevar a cabo el test de Fisher, dado que los tamaños muestrales son pequeños:

```
> Eficacia = matrix(c(5, 10, 11, 9), nrow = 2, dimnames = list(Pulgon = c("Trigo",
     "Cebada"), Muerto = c("Sí", "No")))
> Eficacia

        Muerto
Pulgon   Sí No
  Trigo   5 11
  Cebada 10  9

> fisher.test(Eficacia)

        Fisher's Exact Test for Count Data

data:  Eficacia
p-value = 0.3064
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.079903 1.979949
sample estimates:
odds ratio
   0.41993
```

El p-valor (0.3064) indica que no existe evidencia suficiente de haya diferencia en la eficacia sobre las dos especies. Por tanto, la diferencia observada en la muestra es atribuíble simplemente al azar.

12. A partir de la experiencia descrita en el problema anterior se piensa que cabe la posibilidad de que el insecticida sea más efectivo para el pulgón de la cebada que para el del trigo. Para tratar de confirmar la evidencia a favor de esta hipótesis, se realiza una nueva experiencia con doce ejemplares de cada clase, eligiendo como hipótesis nula que el insecticida es más

efectivo sobre el pulgón del trigo. Transcurridas 24 horas han muerto 9 pulgones de la cebada y 7 del trigo. ¿Qué puede concluirse de este resultado?

Solución:
El contraste a resolver es ahora de la forma:

$$\begin{cases} H_0: & \pi_{Trigo} \geq \pi_{Cebada} \\ H_1: & \pi_{Trigo} < \pi_{Cebada} \end{cases}$$

Nuestras estimaciones a partir de los datos aportados por esta nueva muestra son:

$$p_{Trigo} = \frac{7}{12} = 0{,}58333 \qquad p_{Cebada} = \frac{9}{12} = 0{,}75$$

Como ya ocurría en el problema anterior, el insecticida ha resultado más efectivo sobre la muestra de pulgones de la cebada que sobre la del trigo. Para determinar si esta evidencia es suficiente para poder generalizarla a las poblaciones respectivas debemos llevar a cabo, como antes, el test de Fisher, dado que los tamaños muestrales son pequeños:

```
> Eficacia = matrix(c(7, 9, 5, 3), nrow = 2, dimnames = list(Pulgon = c("Trigo",
     "Cebada"), Muerto = c("Sí", "No")))
> Eficacia

        Muerto
Pulgon   Sí No
  Trigo   7  5
  Cebada  9  3

> fisher.test(Eficacia, alternative = "less")

        Fisher's Exact Test for Count Data

data:  Eficacia
p-value = 0.3334
alternative hypothesis: true odds ratio is less than 1
95 percent confidence interval:
 0.0000 2.7158
sample estimates:
odds ratio
   0.48195
```

El p-valor (0.3334) indica que no existe evidencia suficiente de que la eficacia sobre el pulgón de la cebada sea mayor que sobre el trigo. Por tanto, la diferencia observada en la muestra es atribuíble simplemente al azar.

Si las dos muestras (la de este problema y la del anterior) se han obtenido en las mismas condiciones experimentales, podemos combinar los resultados y llevar a cabo el test con un tamaño de muestra mayor (no podría hacerse si las condiciones experimentales fuesen distintas, porque en tal caso no sería posible decidir si el resultado obtenido se debe a la eficacia del insecticida o al efecto de las distintas condiciones experimentales). Si combinamos los resultados, tendríamos un total de 12 pulgones del trigo muertos entre 28 utilizados en el experimento, y 19 de la cebada muertos del total de 31 utilizados. Las estimaciones serían ahora:

$$p_{Trigo} = \frac{12}{28} = 25{,}429 \qquad p_{Cebada} = \frac{19}{31} = 0{,}6129$$

y el contraste se realizaría mediante:

```
> Eficacia = matrix(c(12, 19, 16, 12), nrow = 2, dimnames = list(Pulgon = c("Trigo",
     "Cebada"), Muerto = c("Sí", "No")))
> Eficacia

        Muerto
Pulgon   Sí No
  Trigo  12 16
  Cebada 19 12

> fisher.test(Eficacia, alternative = "less")

        Fisher's Exact Test for Count Data

data:  Eficacia
p-value = 0.124
alternative hypothesis: true odds ratio is less than 1
95 percent confidence interval:
 0.0000 1.2849
sample estimates:
odds ratio
   0.47987
```

Aún con la evidencia acumulada por todos los datos, el p-valor (0.124) indica que tal evidencia sigue siendo insuficiente, y los datos no permiten asegurar al 5 % de significación que la eficacia sea mayor en el pulgón del trigo que en el de la cebada.