Handout 13

1. An individual's present route to work (route A) results in 40 minutes of travel time per trip. An alternative (route B) has been suggested by a friend who claims that will reduce on average the travel time. Suppose that the new route was tried on 10 randomly chosen occasions with the following times resulting:

   44, 38.5, 37.5, 39, 38.2, 36, 42, 36.5, 36, 34

   Assuming normality, do these data establish the claim that the new route is shorter, at the: a) 1%, b) 5% and c) 10% level of significance?

   (Adapted from problem 5, p. 438, [1])

2. (Refer to the previous problem for background.) Yet another friend suggests route C, which he claims is better. Now you want to compare route B and route C. You test route C eight times, with measured times:

   35, 44.5, 33.4, 37.2, 47, 36.2, 41.1, 33.1

   Assuming normality and equal variances, do these data establish the claim that route C is shorter on average than route B at the 5% significance level?

3. Even if on average route C turns out to be faster than route B, there is concern that sometimes, because of traffic jams or otherwise, it may take considerably longer: that forces one to leave for work earlier, in anticipation of possible problems.

   One way of checking if route C is less reliable than route B as regards trip times would be to test for equal variances, versus the alternative that the variance of travel time is greater with route C. Do that test and report the p-value of the result.

4. What are the implications of your answer to problem 3 on your solution to problem 2?

5. You become concerned that normality is not a plausible assumption in problem 2 and decide to perform a permutation test of $H_0 : m_B = m_C$ versus $H_a : m_B > m_C$.

   (a) Describe what you would do.

   (b) Do it. (For this you need of course to use a computer.)

6. A question of medical interest is whether regular running helps lower the blood pressure. To help decide the issue, 16 persons, 8 runners and 8 non-runners are measured with results that follow:

Table 1: Blood pressure of 16 persons, runners and non runners

	Subject							
	1	2	3	4	5	6	7	8
Non-runners	134	122	118	130	144	125	127	133
Runners	130	120	123	127	128	121	120	125

(a) Perform a two-population t-test, assuming normality and equal variances.

(b) Criticize your method.

(c) Consider that instead of 8 runners and 8 non-runners, you take only 8 non-runners and measure their pressure before and after a running program. Assume you obtain,

	Subject							
	1	2	3	4	5	6	7	8
Before	134	122	118	130	144	125	127	133
After	130	120	123	127	128	121	120	125

(Notice: exactly the same data, but the design of the experiment is quite different.) Explain why you would *not* perform a two-population t-test to assess whether the running program had any effect.

(d) Perform a paired t-test to see if the program appears to have had any effect.

References

[1] Sheldon M. Ross. *Introductory Statistics, Third Edition.* Academic Press, 2010.

Respuestas

1. Tenemos que ver si la nueva ruta supone reducción significativa sobre los 40 minutos que toma la establecida, A; es decir, si hay evidencia suficiente para concluir que el tiempo promedio con la nueva ruta, B, es inferior a 40 minutos.

   Lo podemos plantear como un contraste de $H_0 : m = 40$ frente a la alternativa $H_a : m < 40$. Bajo el supuesto de normalidad se trata de un contraste standard: utilizaremos el estadístico:

   $$T = \frac{\overline{X} - 40}{s}\sqrt{n-1}$$

   que bajo H_0 se distribuirá como una t_{n-1}. Realizando los cálculos,

   ```
   > x  <- c(44, 38.5, 37.5, 39, 38.2, 36, 42, 36.5, 36, 34)
   > nx <- length(x)
   > xbar <- mean(x)
   > sx2   <-   ( sum( (x-xbar)^2) ) / nx
   > sx2

   [1] 7.9501

   > T    <- ( (xbar - 40)  / sqrt(sx2) ) * sqrt(nx-1)
   > T

   [1] -1.94709
   ```

   Este valor del estadístico T lo hemos de comparar con los cuantiles dejando *a su izquierda* colas del 10%, 5% ó 1%. La región crítica esta formada por valores pequeños de T, porque son los que arrojan evidencia contra la hipótesis nula:

   ```
   > qt( c(0.10, 0.05, 0.01), df=9)

   [1] -1.383029 -1.833113 -2.821438
   ```

   Como -1.94709 es menor que -1.383029 y -1.833113, rechazaríamos la hipótesis nula (es decir, declararíamos la nueva ruta como significativamente más corta) a los niveles de significación del 10% y 5%. La evidencia no sería suficiente como para rechazar H_0 al nivel de significación del 1%.

   Hay que señalar que para emplear el estadístico que se ha empleado el estimador de la varianza ha de ser el utilizado (sesgado, máximo-verosímil, utilizando como divisor n y no $n-1$). Si hubiéramos empleado la función `var` para calcular la varianza,

```
> var(x)

[1] 8.833444
```

tendríamos que ajustar el resultado para que coincidiera con el nuestro:

```
> var(x) * (nx-1) / nx

[1] 7.9501
```

Si en lugar de considerar unos niveles de significación pre-establecidos quisiéramos calcular el p-valor asociado al estadístico haríamos:

```
> pt(T, df=nx-1)

[1] 0.04167765
```

Una vez que sepan hacerlo manualmente, puede introducirse la función `t.test` que hace todo por ellos:

```
> t.test(x, alternative="less", mu=40)

        One Sample t-test

data:  x
t = -1.9471, df = 9, p-value = 0.04168
alternative hypothesis: true mean is less than 40
95 percent confidence interval:
     -Inf 39.89288
sample estimates:
mean of x
    38.17
```

2. Ahora hemos de comparar las rutas B y C, suponiendo normalidad e iguales varianzas. El contraste es de nuevo unilateral, pues C se presenta como mejor opción que B: $H_0 : m_B = m_C$ versus $H_a : m_B > m_C$. Tenemos ya calculadas la media y varianza muestral correspondientes a la ruta B. Hacemos los cálculos correspondientes para C:

```
> y    <- c(35, 44.5, 33.4, 37.2, 47, 36.2, 41.1, 33.1)
> ny   <- length(y)
> ybar <- mean(y)
> sy2  <- ( sum( (y-ybar)^2) ) / ny
```

El estadístico es ahora:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{s^2 \left[\frac{1}{n_x} + \frac{1}{n_y} \right]}}$$

siendo s^2 el estimador de la varianza, supuesta común, que hace uso de información en ambas muestras:

$$s^2 = \frac{n_x s_x^2 + n_y s_y^2}{n_x + n_y - 2}.$$

(Notemos que n_x y n_y no son iguales, ni es preciso que lo sean.) La región crítica estará situada a la derecha, pues son valores de $\overline{X}$ mayores que los de $\overline{Y}$ los que arrojan evidencia contra la hipótesis nula y a favor de la alternativa.

```
> nx

[1] 10

> ny

[1] 8
```

Realizando los cálculos,

```
> T <- (xbar - ybar) /
+       ( sqrt(1/nx + 1/ny) * sqrt( (nx*sx2 + ny*sy2) / (nx + ny - 2)) )
> T

[1] -0.1375667
```

El valor del estadístico es negativo, significando que el tiempo medio de viaje utilizando la ruta C (`ybar`) es mayor que utilizando la ruta B (`xbar`): claramente no rechazaremos la hipótesis de que ambas tiene un tiempo medio de viaje igual en beneficio de la alternativa. De todos modos, el cálculo del p-valor sería:

```
> 1 - pt(T, df = nx + ny - 2)

[1] 0.5538504
```

(Nótese el diferente modo de calcular el p-value respecto al caso precedente; ahora la región crítica esta formada por la cola derecha de la distribución del estadístico.)

Una vez que saben hacer los cálculos, pueden recurrir a utilizar `t.test`:

```
> t.test(x=x, y=y, alternative="greater", var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = -0.13757, df = 16, p-value = 0.5539
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -3.662389       Inf
sample estimates:
mean of x mean of y
  38.1700   38.4375
```

3. Lo que se plantea ahora es un contraste de igualdad de varianzas. Bajo la hipótesis nula es: $H_0 : \sigma_x^2 = \sigma_y^2$. La alternativa es $H_0 : \sigma_y^2 > \sigma_x^2$ (ruta C con mayor incertidumbre).

   El estadístico de contraste será:

$$T = \frac{n_x s_x^2/(n_x - 1)}{n_y s_y^2/(n_y - 1)}$$

   que bajo la hipótesis nula sigue una distribución $\mathcal{F}_{n_x-1,n_y-1}$. Si la alternativa es cierta, debemos esperar valores mayores de s_y^2 de los que se presentarían bajo H_0 y por tanto un valor de T menor: la región crítica estará formada por la cola izquierda.

```
> T <- (nx*sx2 / (nx-1)) / (ny*sy2 / (ny-1))
> T

[1] 0.3265103

> pf(T,nx-1,ny-1)

[1] 0.06064743
```

   Una vez lo saben hacer, pueden utilizar:

```
> var.test(x=x, y=y, ratio=1, alternative="less")

        F test to compare two variances

data:  x and y
F = 0.32651, num df = 9, denom df = 7, p-value = 0.06065
alternative hypothesis: true ratio of variances is less than 1
95 percent confidence interval:
 0.000000 1.075115
```

```
sample estimates:
ratio of variances
         0.3265103
```

4. No hay evidencia fuerte de que las varianzas sean diferentes, lo que comprometería la validez del contraste de igualdad de medias realizado precedentemente.

5. Si la hipótesis de normalidad nos pareciera sospechosa o insostenible, podríamos recurrir a hacer un contraste de permutación. La diferencia de medias observada es:

```
> T <- mean(x) - mean(y)
> T

[1] -0.2675
```

Si realmente las rutas tuvieran ambas la misma duración, dicho valor de T debería ser uno esperable al hacer la diferencia de medias de $n_x = 10$ y $n_y = 8$ valores obtenidos por muestreo aleatorio del conjunto de los 18 valores observados. La alternativa es que la ruta C es más corta: de ser cierta, el valor de T sería mayor de lo esperable al hacer dicho muestreo aleatorio.

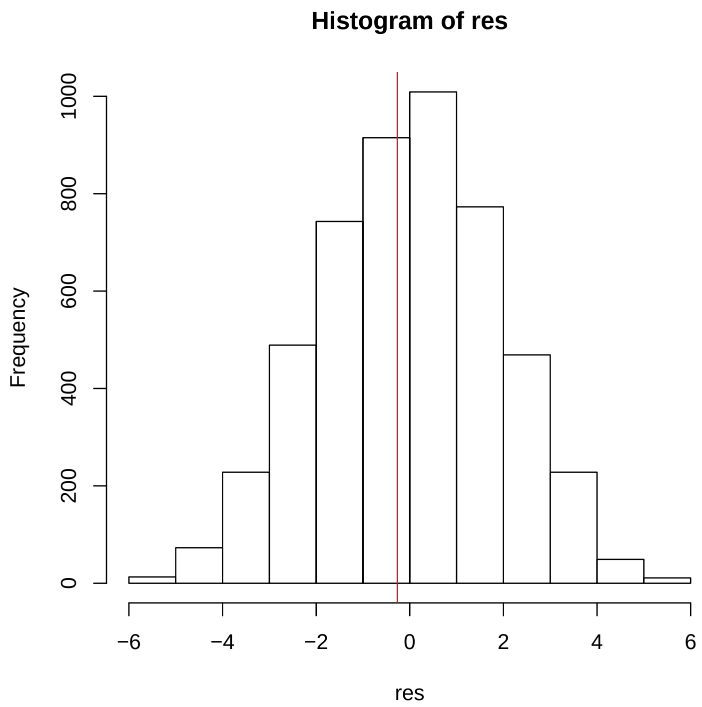
Repitamos esta operación de muestreo para ver qué valores son esperables:

```
> N   <- 5000             # Número de muestreos
> res <- rep(0,N)         # Para alojar los resultados
> S   <- c(x,y)           # Muestra conjunta
> p   <- length(S)        # Tamaño de la misma
> for (i in 1:N) {        # N replicaciones calculando cada vez
+      j   <- sample(p,p)  # el análogo de T
+      tmp <- S[j]
+      res[i] <- mean( tmp[1:nx] ) - mean( tmp[(nx+1):p] )
+      }
```

Veamos ahora cual es la posición del T observado en comparación con lo esperable bajo H_0 y calculemos el nivel de significación aproximado:

```
> hist(res)
> abline(v=T, col="red")
> pvalue <- sum(res > T) / N
> pvalue

[1] 0.5512
```



No hay razón pues para rechazar H_0.

6. (a) El contraste de igualdad de medias no tiene dificultad. Puede hacerse manualmente con ayuda de los resultados intermedios que se facilitan (medias y varianzas estimadas) o recurrirse a la función `t.test` de R:

```
> non.runners <- c(134, 122, 118, 130, 144, 125, 127, 133)
> runners     <- c(130, 120, 123, 127, 128, 121, 120, 125)
> mean(runners)                        #  Para resolución manual

[1] 124.25

> mean(non.runners)

[1] 129.125

> var(runners) * (7 / 8)

[1] 12.9375

> var(non.runners) * (7 / 8)

[1] 57.10938

> #
> t.test(x=runners, y=non.runners, alternative="less",
+        var.equal=TRUE)
```

```
        Two Sample t-test

data:  runners and non.runners
t = -1.5411, df = 14, p-value = 0.07279
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf 0.6966163
sample estimates:
mean of x mean of y
  124.250   129.125
```

No se rechaza la hipótesis nula de igualdad.

(b) El problema con este enfoque es que tiene muy poca potencia: estamos intentando detectar una diferencia pequeña entre grupos en presencia de una gran variabilidad dentro de cada grupo (la tensión arterial de persona a persona varía mucho en comparación con la mejora que cabría atribuir al hecho de correr).

(c) Si, por el contrario, examinamos a *las mismas* personas antes y después de un programa de entrenamiento, podemos calcular la diferencia para cada persona antes y después y contrastar si dicha diferencia tiene media cero (ninguna mejora) frente a la alternativa de media negativa (se ha reducido la presión arterial). Esto cancela el "efecto persona" y nos permite concentrarnos en el "efecto correr" —que es lo que realmente nos interesa—.

(d) Un test de datos pareados (o "comparaciones emparejadas") adopta la siguiente forma:

```
> antes   <-  c(134, 122, 118, 130, 144, 125, 127, 133)
> despues <-  c(130, 120, 123, 127, 128, 121, 120, 125)
> dif     <- despues - antes
> dif

[1]  -4  -2   5  -3 -16  -4  -7  -8

> (mdif  <- mean(dif))

[1] -4.875

> (s2dif <- var(dif) * (7 / 8))

[1] 31.10938
```

Para contrastar la hipótesis de que la media de las diferencia es cero frente a la alternativa de que es menor que cero, procedemos como en el primer problema, calculando el estadístico:

$$T = \frac{\overline{X} - 0}{s}\sqrt{n-1}$$

El valor es:

```
> T <- ( (mdif - 0) / sqrt(s2dif) ) * sqrt(8 - 1)
> T
[1] -2.31248
```

y la distribución bajo la hipótesis nula es t_7. El nivel de significación empírico o p-valor es:

```
> pt(T, df=7)
[1] 0.02699459
```

A diferencia de lo que ocurría antes, encontramos ahora evidencia bastante fuerte de que el programa de entrenamiento fue eficaz.

Una vez que saben hacerlo manualmente, no es preciso que lo vuelvan a hacer nunca más. Pueden invocar `t.test` así:

```
> t.test(x=despues, y=antes,
+        alternative="less", paired=TRUE)
        Paired t-test

data:  despues and antes
t = -2.3125, df = 7, p-value = 0.02699
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -0.880989
sample estimates:
mean of the differences
                 -4.875
```

Orientaciones docentes

1. Un contraste ordinario de diferencia de medias en el caso de ocho personas examinadas antes y después de correr no sería correcto, porque se violaría el supuesto de independencia entre los valores de ambas muestras (son las mismas personas, antes y después: hay que suponer relación entre sus valores).

2. No preguntamos, *per se* cosas sobre contrastes emparejados: sólo se introducen como contrapunto para que entiendan el requisito de valores independientes en las dos muestras.