3.7 Pruebas para muestras apareadas

La utilización de muestras apareadas es una situación común en la práctica de investigación.

Esta estrategia de investigación surge cuando cada observación para un tratamiento está apareada con otra observación para el otro tratamiento.

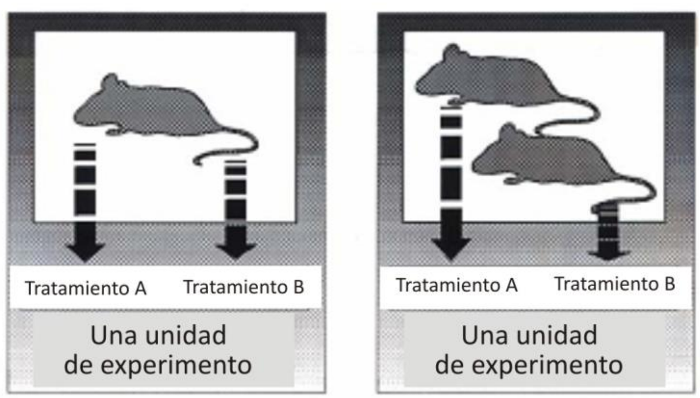
Esta pareja está compuesta por las mismas unidades experimentales observadas dos veces en distintos momentos de la investigación o por dos unidades experimentales con experiencias comunes (muy similares).

El procedimiento consiste en buscar pares de unidades experimentales con características similares y asignar aleatoriamente cada unidad del par a cada uno de los dos tratamientos en estudio.

Por ejemplo, es muy común la situación en que se observa a una misma unidad experimental midiendo alguna característica antes y después de aplicar un cierto tratamiento.

Un par puede estar compuesto, por ejemplo, por dos ratas machos de la misma camada, dos pacientes de la misma edad y sexo, dos mitades de una hoja, dos productos industriales fabricados por la misma máquina y el mismo operario, etc.

El objetivo del apareamiento consiste en hacer más precisa la comparación entre los tratamientos, teniendo en cuenta que los miembros del par sean lo más parecido posible y que la única diferencia que se encuentre en las mediciones efectuadas sobre ellos se deba exclusivamente al tratamiento en estudio.



De la misma manera que con muestras independientes, podemos simbolizar cada medición con x_{ij}, donde el subíndice i refleja la unidad experimental y el subíndice j el tratamiento.

Cada una de las mediciones representa:

- el efecto del tratamiento (t_j),
- el efecto propio de la unidad experimental que es común a ambos tratamientos (p_i),
- el efecto de la variabilidad intrínseca de cada unidad experimental (e_i).

Luego, podemos expresar matemáticamente esta situación del siguiente modo:

$$X_i = t_j + p_i + e_i \quad i = 1,2,\ldots,n \quad j = 1,2$$

La diferencia que existe con el modelo estructural planteado en el caso de muestras independientes, es la presencia del término p_i. En el caso de muestras independientes este término estaba presente pero estaba confundido con el error experimental y no podía distinguirse por sí mismo.

De manera similar al modelo estructural planteado en el caso de muestras independientes, este modelo también lleva implícitos algunos supuestos que veremos a continuación.

Supuestos del Modelo

- **Primer supuesto**

Los errores e_i se distribuyen normalmente con media 0 y varianza σ^2_θ.

$$e_i \sim N(0, \sigma_\theta^2)$$

- **Segundo supuesto**

Las variables t_j y e_i no están correlacionadas.

Estos dos supuestos ya han sido explicados cuando tratamos el tema correspondiente a dos muestras independientes.

El tercer supuesto de igualdad de varianzas poblacionales no es un requisito en la aplicación de la metodología inferencial cuando se consideran muestras dependientes dado que ahora trabajaremos con una variable a la cual denominaremos diferencia y simbolizaremos con d_i.

A continuación hablaremos del modelo estadístico correspondiente a ensayos apareados, comentaremos el diseño experimental y desarrollaremos la prueba de hipótesis adecuada.

Modelo Estadístico

Supongamos que queremos comparar el efecto producido por dos tratamientos A y B.

Vamos a llamar x_{ia} a la respuesta de la i-ésima unidad al tratamiento A.

Luego, cada observación va a ser el resultado de la siguiente suma:

$$X_{ia} = t_a + p_i + e_{ia}$$

donde:

x_{ia} = respuesta de la i-ésima unidad experimental al tratamiento A.

t_a = efecto del tratamiento.

p_i = efecto común asociado a la i-ésima unidad experimental.
e_{ia} = error experimental no controlado por el investigado.

Si consideramos una unidad experimental expuesta al tratamiento B, tenemos:

$$X_{ib} = t_b + p_i + e_{ib}$$

Vamos a introducir ahora una nueva variable a la que llamaremos d_i.
d_i será la diferencia entre la medición de la respuesta de la unidad experimental al tratamiento A y la medición de la respuesta de la misma unidad cuando recibe el tratamiento B.

En símbolos:

$$d_i = x_{ia} - x_{ib} = (t_a + p_i + e_{ia}) - (t_b + p_i + e_{ib}) = (t_a - t_b) + (e_{ia} - e_{ib})$$

Vemos que al efectuar la diferencia entre dos observaciones apareadas, ha desaparecido el término p_i. Esto implica que se ha eliminado el efecto común asociado con el individuo i-ésimo.

Entonces, cuando se trabaja con muestras apareadas conviene utilizar como variable la diferencia de las respuestas de los individuos de cada par. La diferencia d_i refleja el efecto del tratamiento sobre el par y no sobre cada uno de ellos por separado.

La variable d_i depende de la diferencia de los efectos de cada uno de los dos tratamientos considerados más la diferencia de los correspondientes términos de error.

El término t_a - t_b es constante para todas las unidades experimentales y el término e_{ia} - e_{ib} varía de unidad experimental en unidad experimental.

Por este motivo, toda la variabilidad que puede estar presente en el ensayo se deberá exclusivamente al error experimental.

De acuerdo al modelo estadístico que hemos establecido, podemos calcular el promedio correspondiente a la variable d_i que simbolizaremos como d_i.

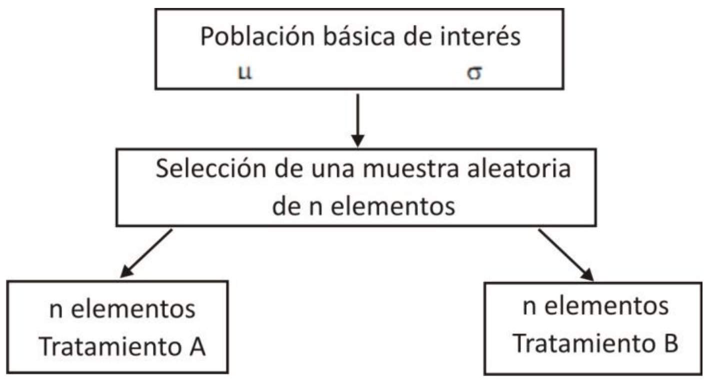
$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{\sum_{i=1}^{n} (t_a - t_b)}{n} + \frac{\sum_{i=1}^{n} (e_a + e_b)}{n}$$
$$= n \cdot \frac{(t_a - t_b)}{n} = (\bar{e}_a - \bar{e}_b) = (t_a - t_b) + (\bar{e}_a - \bar{e}_b)$$

siendo:

$$\bar{e}_a = \frac{\sum_{i-1}^{n} \bar{e}_a}{n} \quad y \quad \bar{e}_b = \frac{\sum_{i-1}^{n} \bar{e}_b}{n}$$

Diseño Experimental

Veremos a continuación el diseño experimental correspondiente a la situación en que se utilizan muestras apareadas.



Cuando se utiliza este tipo de diseño, el investigador selecciona aleatoriamente una muestra de n individuos a los cuales se observa bajo dos situaciones de experimentación. Los experimentos antes-después son un caso particular de esta situación de experimentación.

También pueden seleccionarse aleatoriamente n pares de individuos donde los individuos de cada par presentan características muy similares.

Son muy conocidos los experimentos donde se investigan comportamientos sicológicos de hermanos gemelos, los estudios retrospectivos en los que se aparean individuos similares donde la única característica que distingue a los individuos de un par es, por ejemplo, la presencia de alguna patología, etc.

En todos los casos, el interés del investigador radica en establecer si ha habido una diferencia significativa cuando se aplicaron los dos tratamientos en estudio o entre las dos situaciones de experimentación (antes después).

Para ello deberá recurrir a una prueba de hipótesis, tema que veremos a continuación.

Prueba de Hipótesis

Trataremos de dilucidar estas cuestiones teóricas por medio de un ejemplo.

Supongamos que en un ensayo clínico se está probando la eficacia de una droga para disminuir el nivel de la presión arterial.

El experimento consiste en seleccionar aleatoriamente un grupo de pacientes hipertensos, medirles la presión arterial y registrarla.

A continuación se les administra la droga en cuestión y luego de un tiempo razonable se vuelve a medir la presión.

Este es un caso típico de un experimento **antes-después** pues el mismo individuo es observado antes y después de administrar el tratamiento en estudio.

Como primera medida se deberá establecer la hipótesis a ser probada.

En general, la hipótesis es la siguiente:

$$H_0)\ \mu_A = \mu_B \quad ó \quad \mu_A - \mu_B = 0$$

La hipótesis nula estaría indicando que la droga no ha modificado los promedios de la presión sanguínea luego de ser administrada.

Una hipótesis alternativa podría ser:

$$H_1)\ \mu_A \neq \mu_B \quad ó \quad \mu_A - \mu_B \neq 0$$

Calculamos ahora el promedio de la variable diferencia, estadístico que utilizaremos para efectuar la prueba:

$$\bar{d} = (\bar{x}_a - \bar{x}_b)$$

La varianza será:

$$S^2_{di} = S^2\ (x_{ia} - x_{ib}) = S^2(x_{ia}) + S^2(x_{ib}) - 2\ COV\ (x_{ia}, x_{ib})$$
$$= S^2(x_{ia}) + S^2(x_{ib}) - 2\ r_{ab} \cdot S_a \cdot S_b$$

donde

$$COV\ (x_{ia}, x_{ib}) = r_{ab} \cdot S_a \cdot S_b$$

$$\left[\text{Recordar que} \quad r_{ab} = \frac{COV\ (x_{ia}, x_{ib})}{S_a \cdot S_b} \right]$$

siendo r el coeficiente de correlación lineal que existe entre las mediciones.

Una de las ventajas primordiales de los diseños de mediciones repetidas es la reducción potencial de la varianza debida al error experimental.

Para este diseño:

$$S_{\bar{d}}^2 = S^2\,(\bar{x}_a - \bar{x}_b) = \frac{S_d^2}{n} = \frac{S_a^2}{n} + \frac{S_b^2}{n} - 2\,r_{ab}\,\frac{S_a \cdot S_b}{n}$$

La correspondiente estimación para el caso de observaciones no correlacionadas (muestras independientes) es:

$$S^2\,(\bar{x}_a - \bar{x}_b) = \frac{S_a^2}{n} + \frac{S_b^2}{n}$$

Entonces, si la correlación entre las observaciones es positiva, la estimación del error experimental de un diseño de comparaciones apareadas donde se emplea S^2_d será menor que la obtenida de un diseño de observaciones no correlacionadas en el factor $2\,r_{ab} \bullet S_a \bullet S_b$.

Luego, si d se distribuye aproximadamente normal, el estadístico:

$$t = \frac{\bar{d} - (\mu_A - \mu_B)}{\sqrt{\frac{S_d^2}{n}}}$$

se distribuye como una t de Student con n - 1 grados de libertad.

Los grados de libertad para este estadístico son (n - 1) mientras que para el caso de muestras independientes son 2 (n - 1). Lo que sucede es que al tomar como variable a la diferencia d_i se pierde información sobre (n - 1) valores.

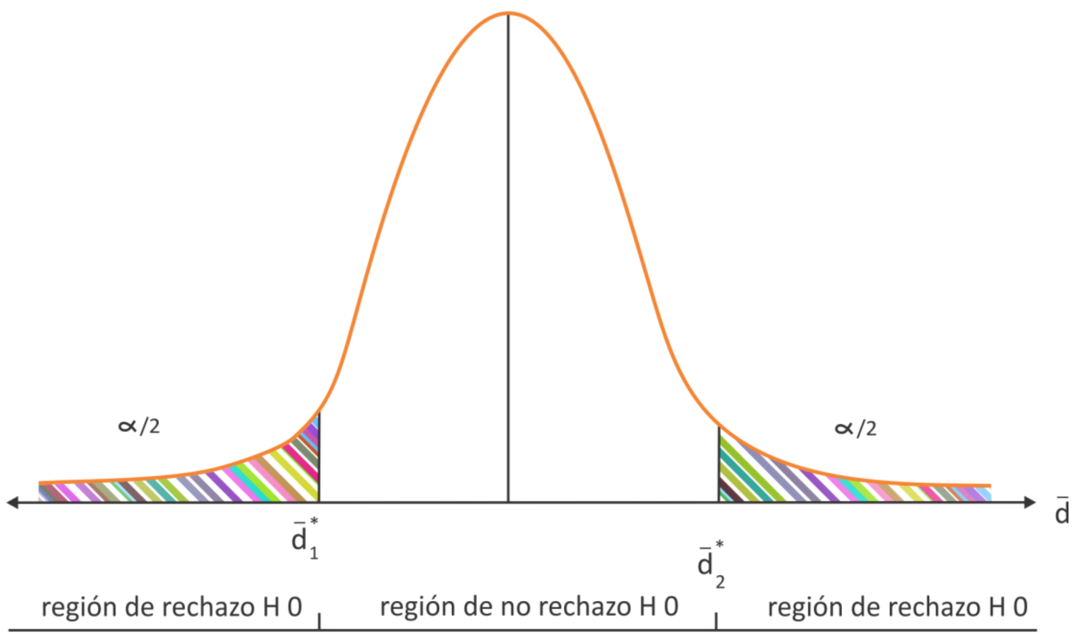
Por tal motivo, antes de considerar que un diseño de mediciones repetidas es más eficiente que otro diseño que no tenga mediciones repetidas, la disminución del error experimental debe ser suficientemente importante como para justificar la reducción de los grados de libertad.

En áreas de la investigación en las cuales la varianza asociada con el término p_i en el modelo de muestras apareadas es grande, el control que introduce el diseño de mediciones repetidas sobre esta fuente de variación incrementa enormemente la sensibilidad del experimento.

Para finalizar el procedimiento de prueba de hipótesis nos falta especificar los valores críticos y las reglas de decisión.

Si la hipótesis es bilateral se deberán calcular dos valores críticos utilizando para ello la distribución t.

Gráficamente, tenemos:



Los valores críticos se calculan como:

$$\bar{d}_1^* = t_{n-1,\left(\frac{\alpha}{2}\right)} \cdot \sqrt{\frac{S_d^2}{n}}$$

$$\bar{d}_2^* = t_{n-1,\left(1-\frac{\alpha}{2}\right)} \cdot \sqrt{\frac{S_d^2}{n}}$$

Las reglas de decisión serán:

Si $\bar{d} < \bar{d}_1^*$ o si $\bar{d} > \bar{d}_2^*$ se rechazará H_0

Si $\bar{d}_1^* < \bar{d} < \bar{d}_2^*$ no se rechazará H_0

Veremos a continuación un ejemplo de aplicación.

Una investigadora educacional desea estudiar la eficacia de un nuevo método para mejorar la ortografía de niños de una determinada edad escolar.

Para ello selecciona aleatoriamente un grupo de 7 niños de edades y antecedentes de aprendizaje similares a quienes les dicta un texto de 100 palabras.

Para cada uno de los niños, registra la cantidad de palabras mal escritas.

Durante todo el año implementa el nuevo método y, al finalizar el año escolar, vuelve a dictar el texto registrando nuevamente la cantidad de palabras con errores.

Suponemos que la variable cantidad de palabras con errores sigue una distribución normal.

Los resultados obtenidos son resumidos en la siguiente tabla:

Niño Nº	Errores antes	Errores después	Diferencia(d_i)
1	6	3	3
2	14	8	6
3	8	4	4
4	4	6	-2
5	16	9	7
6	7	2	5
7	19	12	7

La hipótesis nula planteada es:

$$H_0)\ \mu_{\text{antes}} = \mu_{\text{después}}$$

contra la alternativa:

$$H_1)\ \mu_{\text{antes}} \neq \mu_{\text{después}}$$

Para probar la hipótesis debemos calcular la media y la varianza de la variable diferencia en la muestra: d y S^2_d

$$\bar{d} = \frac{\sum_{i-1}^{n} d_i}{n} = \frac{30}{7} = 4{,}29$$

La media de la variable diferencia también podría haber sido calculada como, x_a- x_b obteniéndose el mismo resultado.

$$S_d^2 = \frac{\sum_{i-1}^{n}(d_i - \bar{d})^2}{n-1} = \frac{(3-4{,}29)^2 + (6-4{,}29)^2}{6} + \cdots + \frac{(5-4{,}29)^2 + (7-4{,}29)^2}{6} = 9{,}90$$

Entonces, la varianza de d es:

$$S_{\bar{d}}^2 = \frac{S_d^2}{n} = \frac{9{,}90}{7} = 1{,}415$$

$$S_{\bar{d}} = \sqrt{1{,}415 = 1{,}190}$$

A continuación debemos buscar el valor de t con n − 1 = 6 grados de libertad.
Si fijamos $\alpha = 0.05$, tendremos:

$$t_{6;\,0{,}975} = 2{,}45$$

Los valores críticos serán:

Si $\bar{d} < -2{,}92$ o si $\bar{d} > 2{,}92$ se rechazará H_0 y

Si $-2{,}92 < \bar{d} < 2{,}92$ no se rechazará H_0

Como d = 4.29 > 2.92 debemos rechazar H_0 y concluir que el método para corregir la ortografía ha sido efectivo pues se observa una disminución significativa en la cantidad de palabras con errores.

Veremos ahora cómo implementamos esta prueba utilizando el programa SPSS. Utilizaremos las sentencias ('comparing groups means'), ('t-test') y ('paired').

La salida de computación es la que figura como Gráfico 19 al final del Módulo.

En la primera parte de la output, o salida, aparece una tabla con medidas descriptivas de las variables consideradas.

Para las dos variables (antes y después) se observa el número de casos (Number of cases), la media aritmética (Mean), la desviación estándar (Standard Deviation) y el error estándar (Standard Error).

La segunda tabla calcula medidas con respecto a la variable diferencia.

Se observa la diferencia media, la desviación estándar y el error estándar de esta variable.

Al final de la tabla se registra el valor de la t de Student (t Value), los grados de libertad (Degrees of Freedom) y el nivel de significación de la prueba (2-Tail Prob).

Otro valor que calcula el programa es el coeficiente de correlación estudiado precedentemente.

Se observa el coeficiente (Corr) y una prueba correspondiente al mismo, cuya hipótesis nula está referida a que el coeficiente de correlación poblacional es 0 contra la alternativa de que es distinto de 0. Recordemos que dijimos que cuando el valor del coeficiente de correlación era cercano a 0 señala que las variables no están correlacionadas.

En nuestro ejemplo en particular, $r = 0.868$, valor bastante cercano a 1.

Existen pruebas para probar la hipótesis planteada con referencia al coeficiente de correlación.

En la salida de computación se da el nivel de significación de uno de estas pruebas ($\alpha = 0.011$). Esto nos está indicando que el coeficiente de correlación es significativamente distinto de 0 y por lo tanto, existe una correlación positiva entre las mediciones antes y después.

Con este resultado podemos expresar que, a pesar de que el método de enseñanza ha sido exitoso, los alumnos que tenían muchos errores de ortografía los siguieron teniendo después de recibir el adiestramiento ortográfico y viceversa.

Es en estas situaciones cuando las pruebas apareadas son más eficientes que las pruebas t para muestras independientes. La existencia de una correlación positiva entre las variables justifica la utilización de una prueba apareada pues se disminuye la variabilidad del ensayo.