3.5 Pruebas para diferencia de medias

Supongamos las siguientes situaciones: un investigador educacional desea comparar dos técnicas de enseñanza; un agrónomo desea evaluar la acción de un insecticida comparándolo con la situación en que no se lo aplica; un médico desea comparar dos terapias para el tratamiento del infarto; un ingeniero desea estudiar la resistencia de un producto proveniente de dos proveedores distintos; un especialista en marketing desea evaluar dos tipos de campañas publicitarias, etc.

La cuestión común a todos los campos de investigación puede resumirse de la siguiente manera: entre dos tratamientos, ¿cuál producirá mejores resultados cuando se aplican sobre una población previamente especificada?

Para responder a esta pregunta seguramente el investigador definirá una población objetivo y a continuación tomará una muestra aleatoria de esta población a partir de la cual evaluará el efecto de los dos tratamientos en estudio que denominaremos A y B.

El investigador educacional tomará una muestra de alumnos que respondan a su población objetivo y los asignará aleatoriamente a dos técnicas de enseñanza, el ingeniero tomará una muestra de productos analizando la resistencia de los mismos en función del proveedor, etc.

Supongamos, por ejemplo, una población definida de la siguiente manera:

"Vacas de una determinada raza lechera de cierta edad".

El objetivo de la investigación consiste en comparar los efectos de dos dietas distintas sobre el rendimiento lechero diario de las vacas en cuestión.

La investigación se lleva a cabo en un campo experimental de la Secretaría de Desarrollo Agropecuario y Rural donde se cuenta con un rodeo de 500 vacas con estas características.

Para efectuar la experiencia, el investigador toma una muestra aleatoria de 20 vacas y asigna también aleatoriamente 10 vacas a cada una de las dietas en estudio. Después de un período considerable, mide la producción lechera diaria de cada vaca con el objetivo de determinar si existe diferencia entre los promedios de producción lechera para los dos grupos considerados.

Generalizando, podemos simbolizar a cada medición con x_{ij}, donde el subíndice i hará referencia a la vaca en particular dentro de su grupo correspondiente y el subíndice j al tipo de tratamiento al que fue sometida.

En este caso en particular, i variará de 1 a 10 pues cada dieta se administró a 10 vacas y j variará de 1 a 2 pues se consideran dos tratamientos (dietas).

Así:

x_{21} = producción diaria de leche de la segunda vaca a la que se le adjudicó la dieta A.

x_{92} = producción diaria de leche de la novena vaca a la que se le adjudicó la dieta B.

Ahora bien, ¿qué se esconde detrás de la medición particular de la producción lechera de una vaca en un día determinado?

Cada una de estas mediciones va a reflejar el efecto de la dieta sobre la producción de las vacas más una componente aleatoria que representa la variabilidad intrínseca de cada vaca (producción anterior, estado físico, genética, etc.).

Vamos a simbolizar con tj al efecto de la dieta y con e_i, a la componente aleatoria aportada por cada vaca en particular.

Luego, podemos expresar matemáticamente el modelo estadístico correspondiente a esta situación experimental, de la siguiente manera:

$x_{ij} = tj + ei \qquad i = 1, 2, ..., n \quad j = 1, 2$

Este modelo no hace más que idealizar la situación correspondiente a la producción de cada vaca en el experimento representada mediante la suma del efecto de la dieta (t_j) más el efecto aportado por la condición propia de cada vaca (e_i).

> Se supone que el efecto de la dieta (tj) actúa uniformemente en todas las vacas que integran una dieta en particular. Así, si la dieta es mala, todas las vacas sometidas a dicha dieta producirán menos leche y viceversa.
> En cambio, el término *ei*, varía de unidad experimental (vaca) a unidad experimental dentro de cada grupo. Además, refleja todo otro tipo de variabilidad que no ha sido controlada por el investigador.

Por ello, el término *ei*, se denomina error experimental y está ligado indirectamente a las condiciones en que se efectúa el experimento.

Debemos aclarar en este punto que la única variable que mide y observa realmente el investigador es x_{ij}, y tanto t_j como e_i, son variables que ayudan a interpretar la aplicación de las técnicas estadísticas de análisis de datos pero nunca van a ser realmente medidas.

Supuestos del modelo

Para efectuar comparaciones entre los efectos de los tratamientos que sean inferencialmente válidas, el modelo debe cumplir ciertas restricciones.

Primer supuesto

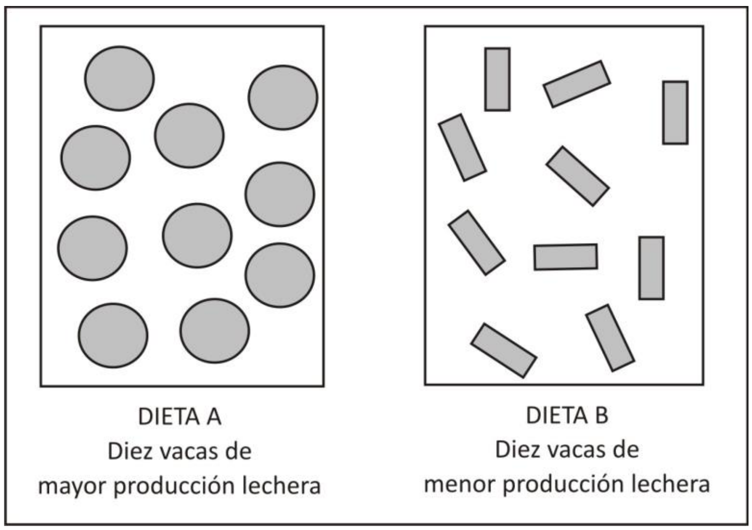
Los errores e_i se distribuyen normalmente con media 0 y varianza σ^2_e .

Segundo supuesto

Las variables t_j y e_i deben ser variables no correlacionadas.

Este supuesto implica que no deben existir errores sistemáticos en la ejecución del experimento. Como vimos en el Módulo anterior, la presencia de estos errores podría deberse a la falta de aleatoriedad en la selección de las unidades experimentales o en la asignación de los tratamientos.

La ausencia de aleatoriedad en la asignación de las vacas a cada una de las dietas podría haber afectado el resultado del experimento. Podría haberse dado la situación extrema en que las 10 vacas de mayor producción lechera (por sus características genéticas) son tratadas con la dieta A y el resto con la dieta B.



Al finalizar la experiencia, el efecto de la dieta estaría confundido con las características genéticas y ambientales de las vacas siendo imposible separarlos. Se ha cometido un error sistemático por falta de aleatorización.

Tercer supuesto

Igualdad de varianzas para los dos grupos en estudio.

Este supuesto implica que las fuentes de variabilidad no controladas por el investigador deben ser comparables en los dos grupos. En otras palabras, las condiciones experimentales deben permanecer constantes durante la ejecución del experimento.

Por ejemplo, las vacas deben permanecer en corrales similares, tener el mismo control sanitario, recibir la misma cantidad de alimento, etc.

Todo lo expresado anteriormente puede sintetizarse en que los dos grupos deben presentar una cierta homogeneidad en cuanto al error experimental.

Ahora bien, ¿cómo se compara estadísticamente el efecto de dos tratamientos?

Para comprender mejor este tema hablaremos en primer lugar de modelo estadístico adecuado al ensayo comparativo de dos grupos, luego nos referiremos al diseño experimental subyacente y por último a la correspondiente prueba de hipótesis.

Modelo estadístico

Si tomamos, por ejemplo, el grupo de 10 vacas asignadas a la dieta A y medimos su rendimiento lechero en un día determinado, podemos calcular el promedio de producción lechera para ese día en particular.

De acuerdo al modelo estadístico preestablecido, calcularemos ahora el promedio de producción lechera para cada uno de los grupos.

$$\bar{x}_a = \frac{\sum_{i=1}^{n_a} x_a}{n_a} = \frac{\sum_{i=1}^{n_a} (t_a + e_a)}{n_a}$$

Pero dijimos que t_a, el efecto de la dieta A, es constante para todas las unidades experimentales que recibieron esa dieta, luego:

$$\bar{x}_a = \frac{n_a t_a}{n_a} + \frac{\sum_{i=1}^{n_a} e_a}{n_a} = (t_a + \bar{e}_a)$$

De la misma manera, la media de la muestra de vacas que recibieron la dieta B puede expresarse como:

$$\bar{X}_b = t_b + \bar{e}_b$$

Pero, el objetivo de un experimento comparativo está dirigido a medir la diferencia entre los efectos de ambos tratamientos.

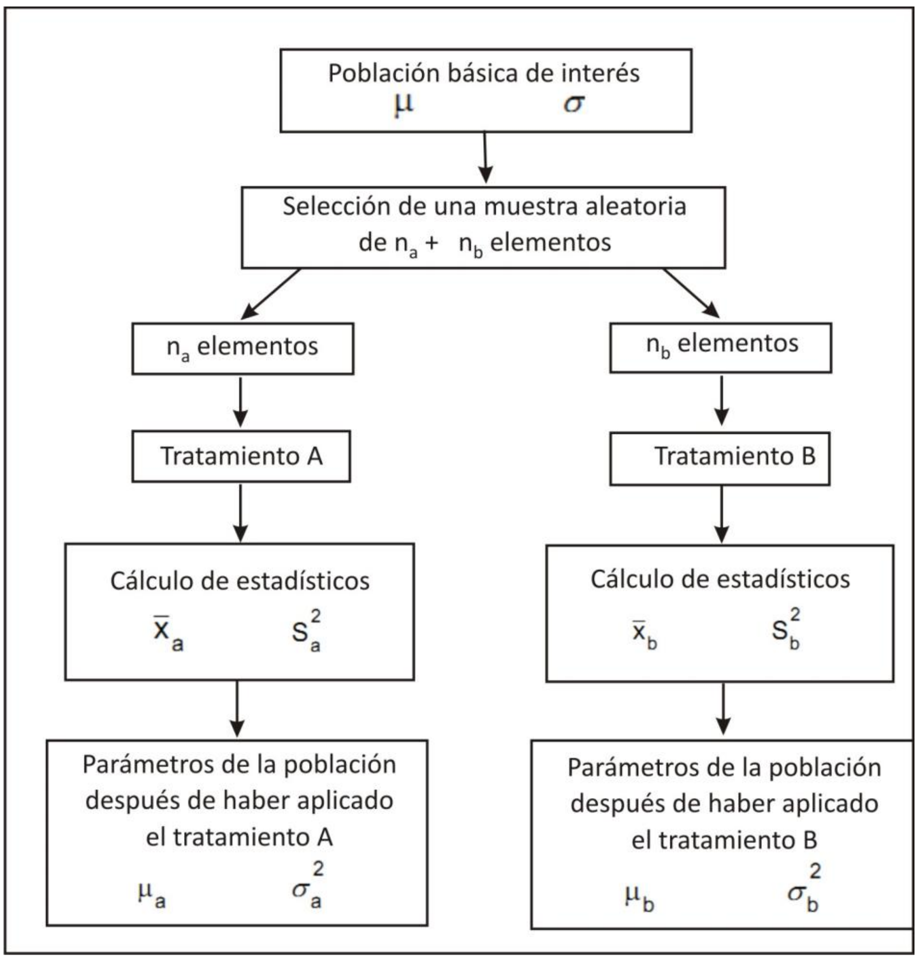
Entonces, si calculamos la diferencia entre los promedios de cada una de las muestras, tendremos:

$$\bar{X}_a - \bar{X}_b = (t_a + t_b) + (\bar{e}_a - \bar{e}_b)$$

Esto nos indica que la diferencia observada entre el promedio de producción lechera del grupo de vacas bajo el tratamiento A y el promedio de producción lechera del grupo de vacas que recibieron el tratamiento B depende, por un lado, de la diferencia entre los efectos de los tratamientos que se prueban y, por otro lado, de las diferencias entre los promedios de los errores experimentales asociados con cada una de las medias de los grupos considerados.

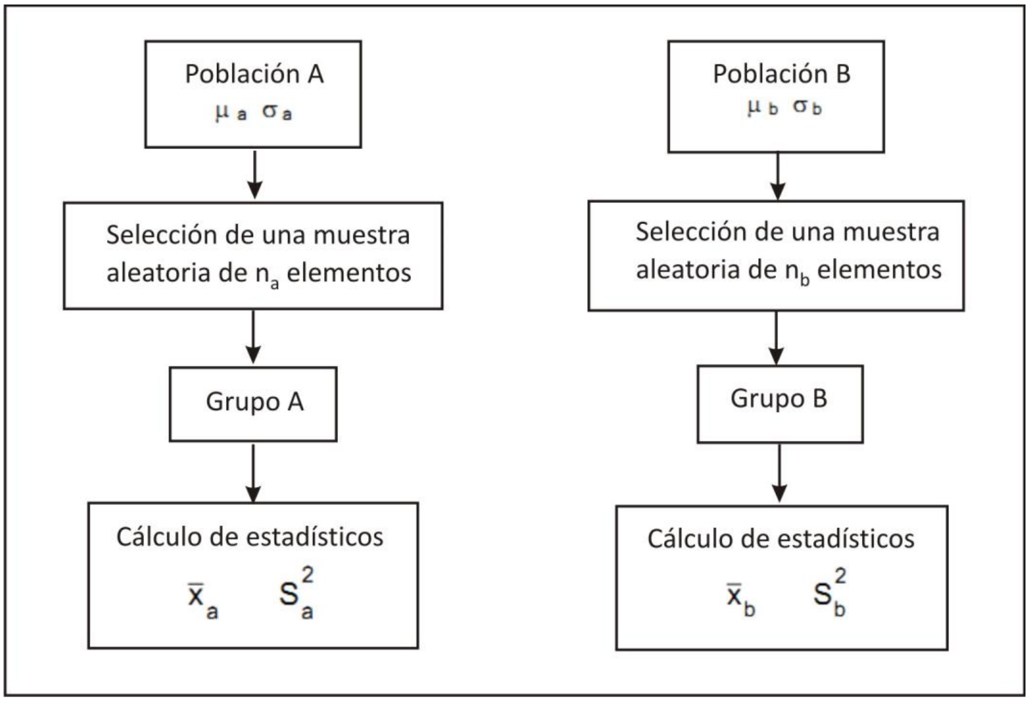
Diseño experimental

Veremos a continuación cómo sería el diseño experimental correspondiente a esta experiencia. El siguiente esquema resume los pasos seguidos por el investigador a través de su experiencia.



Esta situación experimental se presenta cuando se lleva a cabo un experimento controlado. Sin embargo, existen otras situaciones en que el investigador no asigna aleatoriamente las unidades experimentales a los tratamientos sino que parte de dos poblaciones bien diferenciadas donde la aleatoriedad surge, generalmente, en la selección de las muestras. Este tipo de investigación está ligado, principalmente, a los estudios observacionales (estudios retrospectivos, prospectivos, investigación de mercado, etc.).

El esquema que resume esta última situación de investigación es el siguiente:



En ambos casos el interés del investigador se centra en determinar si las medias poblacionales μ_a y μ_b son iguales o no. Para ello deberá recurrir a una prueba de hipótesis.

<u>Prueba de hipótesis</u>

Para efectuar una prueba estadística, como primera medida, se debe establecer la hipótesis que queremos probar. Esta hipótesis generalmente refleja la situación en que los promedios de ambas poblaciones son iguales.

En nuestro ejemplo, la hipótesis nula especifica que las dietas no modifican la producción lechera diaria promedio de las vacas sometidas a ellas.

Como las hipótesis estadísticas se establecen en base a parámetros poblacionales, es lógico pensar que, en este caso particular, una hipótesis nula apropiada, sería:

H_0) $\boldsymbol{\mu_a = \mu_b}$

ó

H_0) $\boldsymbol{\mu_a - \mu_b = 0}$

Si no tenemos ninguna idea previa de cómo actúan las dietas en la producción lechera de las vacas, una hipótesis alternativa conveniente podría responder a una prueba bilateral y ser expresada como:

H$_1$) $\mu_a \neq \mu_b$ ó H$_1$) $\mu_a - \mu_b \neq 0$

La prueba estadístico determinará si la magnitud de esta diferencia puede atribuirse solamente al error experimental o si indica algún efecto mayor que el debido exclusivamente al error experimental.

Si se da la segunda situación planteada, posiblemente diremos que una de las dietas ha producido un aumento en la producción lechera significativamente distinto que la otra.

La evidencia muestral aportada por el investigador para rechazar o no la hipótesis nula planteada será la diferencia de promedios de producción lechera en ambos grupos de vacas considerados: $\bar{x}_a - \bar{x}_b$.

Para determinar la distribución de la variable $(\bar{x}_a - \bar{x}_b)$ vamos a razonar ahora de la misma manera que cuando estudiamos la distribución de la variable aleatoria media muestral. El conocimiento de la distribución de esta variable aleatoria nos permitirá calcular la probabilidad de cometer errores en el proceso inferencial.

Supongamos que repetimos el esquema correspondiente al experimento controlado un número grande de veces. Cada vez empezaríamos con una muestra de $n_a + n_b$ elementos diferentes. En cada repetición obtendríamos una diferencia de medias $(\bar{x}_a - \bar{x}_b)$ y, por lo tanto, podríamos calcular la distribución de esta nueva variable aleatoria: diferencia de medias muestrales.

Si no se cumple el supuesto de normalidad de los e_i por el teorema central del límite podemos decir que, si n_a y n_b son suficientemente grandes, $(\bar{x}_a - \bar{x}_b)$ se distribuirá aproximadamente normal.

La esperanza matemática de esta variable aleatoria será:

$$E(\bar{x}_a - \bar{x}_b) = E(t_a + t_b) + E(\bar{\theta}_a - \bar{\theta}_b) = \mu_a - \mu_b$$

pues

$$E(t_a - t_b) = t_a - t_b = (\mu_a - \mu) - (\mu_b - \mu) = \mu_a - \mu - \mu_b + \mu = \mu_a - \mu_b$$

y

$$E(\bar{\theta}_a - \bar{\theta}_b) = 0 \qquad \text{(Supuesto 1)}$$

y la varianza

$$V(\bar{x}_a - \bar{x}_b) = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}$$

La teoría estadística demuestra que la varianza de la diferencia de variables provenientes de muestras independientes es igual a la suma de las varianzas de cada una de las variables. En nuestro caso,

$$V(\bar{x}_a) = \frac{\sigma_a^2}{n_a} \quad \text{y} \quad V(\bar{x}_b) = \frac{\sigma_b^2}{n_b}$$

Hemos dicho que un supuesto muy importante en este tipo de experimentos comparativos era el de igualdad de las varianzas de ambos grupos de experimentación.

Podemos escribir este supuesto de la siguiente manera:

$$\sigma_a^2 = \sigma_b^2 = \sigma_\theta^2$$

donde σ_θ^2 es el valor de la varianza en la población común a ambos grupos.

Relacionando este concepto con el ejemplo de las vacas que venimos siguiendo,

σ_θ^2 es la varianza de la producción lechera en la población de 500 vacas del campo experimental.

Suponiendo que se cumple este supuesto, entonces,

$$\sigma^2_{(\bar{x}_a - \bar{x}_b)} = \frac{\sigma^2_\theta}{n_a} + \frac{\sigma^2_\theta}{n_b} = \sigma^2_\theta \left(\frac{1}{n_a} + \frac{1}{n_b} \right)$$

Evidentemente, σ_θ^2 es un parámetro poblacional desconocido y entonces debemos estimarlo a partir de las observaciones muestrales.

El mejor estimador muestral de la varianza de error es un promedio ponderado de las varianzas de los dos grupos considerados:

$$S^2_\theta = \frac{(n_a - 1) S^2_a + (n_b - 1) S^2_b}{(n_a - 1) + (n_b - 1)}$$

Luego, la varianza estimada de la diferencia de medias muestrales es:

$$V(\bar{x}_a) = \frac{\sigma^2_a}{n_a} \quad \text{y} \quad V(\bar{x}_b) = \frac{\sigma^2_b}{n_b}$$

Habiendo definido la esperanza y la varianza de la variable diferencia de medias muestrales ($x_a - x_b$), podemos pensar en el siguiente estadístico como criterio de prueba para probar la hipótesis nula establecida:

$$\frac{(\bar{x}_a - \bar{x}_b) - (\mu_a - \mu_b)}{\sqrt{S^2_\theta \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}}$$

Debido a que tenemos dos grupos diferentes que, incluso pueden ser de distinto tamaño muestral, en lugar de dividir la varianza por n se multiplica por la suma de las recíprocas de los tamaños de muestra.

El estadístico, tal como lo hemos presentado, tiene distribución t de Student con $(n_a - 1) + (n_b - 1) = n_a + n_b - 2$ grados de libertad bajo la hipótesis nula.

Utilizaremos entonces la distribución de probabilidad correspondiente a una variable t de Student para calcular el nivel de significación de la prueba y medir, de esta manera, el error que se comete en el proceso inferencial.

Analicemos ahora más en detalle el estadístico planteado. En el numerador aparece la diferencia de las medias muestrales que constituyen la evidencia muestral que posee el investigador para rechazar o no su hipótesis nula. También aparece la diferencia de medias poblacionales (μ_a - μ_b) que no tenemos posibilidad de calcular pero que, si se cumple la hipótesis nula, será igual a 0.

En el denominador tenemos la raíz cuadrada de la varianza ponderada que está calculada enteramente con los datos de la muestra.

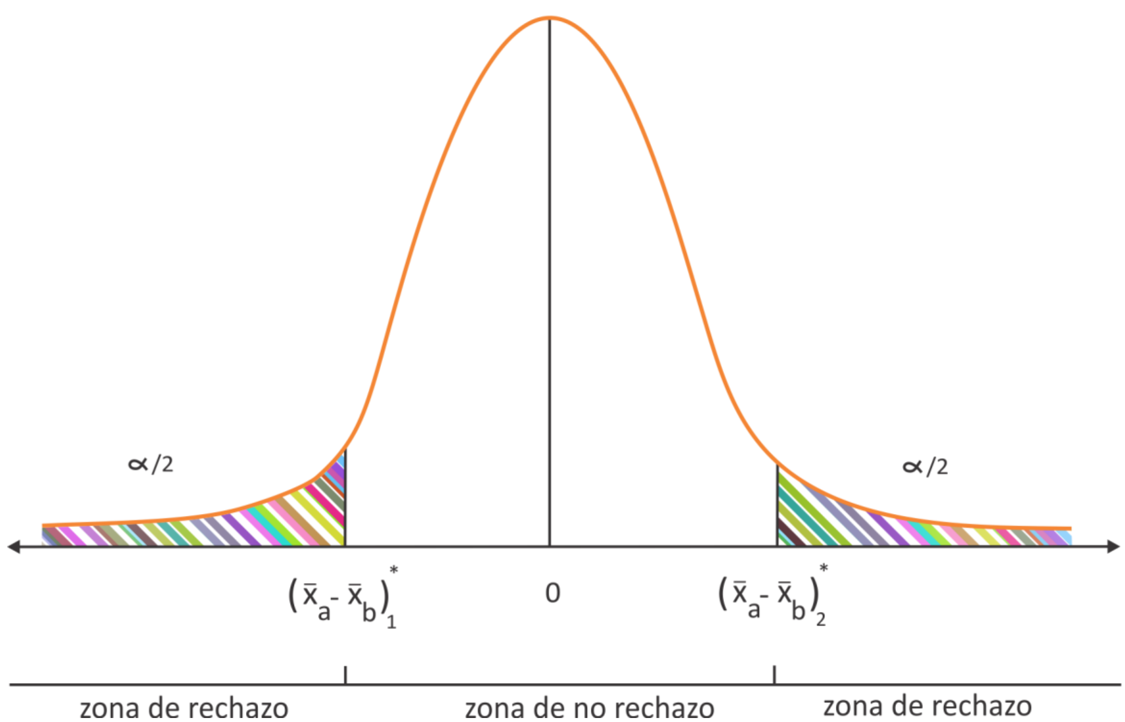
Si se fija de antemano un nivel de significación $\alpha = 0.05$ y se plantea una hipótesis alternativa bilateral, el investigador deberá especificar dos valores críticos, uno a cada lado de la distribución t de Student, de la siguiente manera:

$$(\bar{x}_a - \bar{x}_b)_1 = t_{0,025;(n_a + n_b - 2)} \cdot \sqrt{S_\theta^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}$$

$$(\bar{x}_a - \bar{x}_b)_2 = t_{0,075;(n_a + n_b - 2)} \cdot \sqrt{S_\theta^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}$$

Gráficamente, podemos representar las zonas de aceptación y de rechazo de la hipótesis nula:

Gráfico 1



La regla de decisión será, entonces:

Si $(\bar{x}_a - \bar{x}_b)$ $\quad (\bar{x}_a - \bar{x}_b)^*_1$ ó se rechaza H_0

Si $(\bar{x}_a - \bar{x}_b)$ $\quad (\bar{x}_a - \bar{x}_b)^*_2$

Si $(\bar{x}_a - \bar{x}_b)^*_1$ $\quad (\bar{x}_a - \bar{x}_b)$ $\quad (\bar{x}_a - \bar{x}_b)^*_2$ no se rechaza H_0

La importancia del diseño experimental en el resultado de una prueba estadística.

Para tratar de relacionar aún más la importancia del diseño experimental con el posible resultado de una prueba estadística, volveremos a plantear la siguiente relación:

$$\bar{x}_a - \bar{x}_b = (t_a - t_b) + (\bar{\theta}_a - \bar{\theta}_b)$$

Si las dos dietas influyen prácticamente de la misma manera en la producción lechera, es lógico pensar que el término $(t_a - t_b)$ tenderá a ser 0. En este caso, la diferencia observada por el investigador en los promedios de los respectivos grupos reflejará solamente diferencias entre los errores experimentales. Si el experimento ha sido bien conducido, medirá la variabilidad intrínseca promedio de las vacas de cada grupo.

Entonces, podemos decir que la evidencia muestral $(\bar{x}_a - \bar{x}_b)$ representa solamente diferencias aleatorias o casuales y la prueba seguramente nos llevará a tomar la decisión de no rechazar H_0.

En cambio, si realmente una de las dietas es mejor que la otra para elevar la producción lechera, la diferencia $(t_a - t_b)$ será importante y la evidencia muestral seguramente señalará una diferencia entre promedios de producción lechera que va más allá de una mera diferencia aleatoria.

En este caso, es muy probable que optemos por rechazar H_0 y nuestra conclusión sea que los promedios de producción lechera para ambos grupos no son los mismos.

Ahora bien, ¿qué pasaría si el investigador no conduce correctamente su experimento?

Por ejemplo, puede suceder que el cuidado de las vacas sometidas a la dieta A hubiera quedado a cargo de un empleado menos confiable. Esta persona puede no pesar exactamente la ración, no darse cuenta de la presencia de cierta enfermedad en alguna de las vacas en el tiempo en que dura la experiencia, etc. Es evidente que las vacas del grupo A presentarán un error aleatorio promedio muy superior a las de la dieta B, atendidas, por ejemplo, en forma más homogénea.

Cuando se calcule la diferencia $(\bar{x}_a - \bar{x}_b)$ y ésta aparente ser grande, el investigador puede confundirse pensando que una de las dietas produjo un aumento considerable de la producción lechera cuando en realidad la diferencia está influenciada por una mala conducción del experimento. En este caso particular se introdujo un error sistemático que, al finalizar la experiencia, no puede ser separado del error aleatorio.

Por este motivo, para estar seguros de que la diferencia observada se debe exclusivamente a los efectos de los tratamientos en estudio, la planificación e implementación del experimento juega un rol muy importante.

En síntesis, si las varianzas de las mediciones calculadas en cada uno de los grupos de experimentación no son parecidas, la consideración de la diferencia de las medias muestrales como evidencia para decidir en cuanto a la hipótesis nula planteada no tiene suficiente validez. El efecto de las dietas se encuentra confundido con la variabilidad del material experimental no controlado suficientemente por el investigador.

Volvemos a destacar que esta recomendación es mucho más importante cuando se efectúan estudios observacionales. En este caso no existe un efecto de tratamiento introducido deliberadamente por el investigador sino que el efecto surge de las poblaciones consideradas. El error de muestreo refleja toda las fuentes de variabilidad no controladas por el diseño y las conclusiones del prueba estadística pueden ser erróneas.

En lo referente a heterogeneidad de varianzas aplicaremos los conceptos estudiados al ejemplo del rendimiento lechero de las vacas.

Supongamos que la producción lechera de las 10 vacas sometidas a cada una de las dietas en un día en particular fue la siguiente:

Dieta A	Dieta B
25	22
25	20
28	22
24	23
25	18
27	25
27	23
28	21
25	24
24	21

Con estos datos podemos calcular las medias y las varianzas para cada grupo.

$\bar{x}_a = 25{,}8$ $\quad$ $\bar{x}_b = 21{,}9$

$S_a^2 = 2{,}4$ $\quad$ $S_b^2 = 4{,}1$

El modelo estadístico planteado establece un supuesto muy importante:
la igualdad de las varianzas poblacionales, que será desarrollado en la sección 3.8.