ESTADÍSTICA BÁSICA

Introducción a la Prueba *t* de Student y el Análisis de la Varianza

DESCRIPCIÓN

En este documento se ejemplifica la comparación de medias muestrales mediante el uso de las pruebas estadísticas t de Student y Análisis de Varianza

Prof. Jorge Lorenzo

Estadística básica para Ciencias de la Educación

Foto gentileza de Nick Hillier on Unsplash
https://unsplash.com/@nhillier

Recursos
Educativos
Abiertos

Introducción a la Prueba t de Student y al Análisis de Varianza

Planteo del Problema

En una escuela, se ha enseñado ortografía con el mismo método durante varios años. Algunos docentes han estado trabajando sobre un nuevo método que confían es mejor que el tradicionalmente utilizado. Algunas observaciones informales del rendimiento de los alumnos parecen confirmar que el nuevo método es mejor que el utilizado anteriormente. Deciden realizar un experimento para comprobar que efectivamente el nuevo método corrige la ortografía de los alumnos de una manera más eficiente. Para ello, seleccionan dos muestras aleatorias de alumnos, uno de los cuales recibe el nombre de grupo experimental, el cual quedara conformado por todos los niños que reciben instrucciones con el nuevo método de aprendizaje de la ortografía que se pretende evaluar. El otro recibe el nombre de grupo control, y queda conformado por los niños que serán instruidos con el método tradicional. Las restantes condiciones del experimento se mantienen exactamente iguales para ambos grupos. Es decir, la única diferencia entre ambos grupos es el método de instrucción con el que aprenden la ortografía.

Tras evaluar los resultados con una prueba, los docentes encuentran que, en promedio, el rendimiento de grupo experimental es más alto que el grupo control. Confiando en la eficacia del nuevo método, los docentes plantearon la siguiente hipótesis de trabajo: la diferencia en el promedio obtenido en ambos grupos, se debe a la influencia del tipo de método utilizado para enseñar ortografía a los alumnos. Contrario a este argumento es el supuesto que las diferencias encontradas se debe a variaciones aleatorias, que no han sido influidas por el tipo de método utilizado. La cuestión aquí planteada implica decidir si las diferencias pueden ser atribuidas a un factor conocido (el método de enseñanza de la ortografía), o bien, aceptar que tales diferencias son producto del azar.

Existe una manera de optar por una u otra explicación, basándonos en un modelo estadístico: la distribución *t de Student*. Sobre la base de tal distribución se asienta la conocida prueba t de Student que permite comparar promedios de muestras independientes y determinar si las diferencias encontradas se deben a un factor de influencia. En los párrafos que siguen se explicará en detalle el razonamiento de tal prueba y la aplicación al caso concreto de nuestro ejemplo.

La prueba t de Student

La prueba t de Student fue diseñada por un matemático llamado W. Gosset, se la utiliza cuando la prueba de hipótesis implica una comparación entre dos medias muestrales, siendo las muestras independientes (dos muestras diferentes) o dependientes (una muestra evaluada en dos momentos distintos).

La expresión distribución *t* designa una familia de distribuciones teóricas que sirven a la prueba de hipótesis, cuando las muestras son pequeñas. La forma de la distribución *t* es similar a la curva de distribución normal, solo que más aplanada. Por esto mismo, las características principales de estas curvas es que son unimodales y simétricas, con una media igual a cero. Las distribuciones *t* son una familia de curvas, porque su forma depende del número de casos en la muestra. El número de casos determina los grados de libertad (*gl*) y estos a su vez determinan la forma de la distribución.

Los grados de libertad se refieren al número de valores de una serie estadística que pueden variar libremente después que se ha impuesto ciertas restricciones a la serie de datos. En la prueba *t*, los grados de libertad se representan como N-1 (número de casos menos uno). Como se mencionó, la forma de la distribución *t* es similar a la curva normal, solo que más aplanada en la media y más extendida en los extremos. En la curva de distribución normal, el puntaje crítico de rechazo de la hipótesis nula (H_0), está dado por los valores Z ± 1.96, cuando se ha establecido el error tipo I en un nivel de α=0.05 (en una prueba de dos extremos). En la familia de curvas de la distribución t, estos puntajes críticos para el rechazo de la H_0 varían con los *gl*. Así, cuando los *gl* son iguales a 3, el puntaje que demarca la zona de rechazo estará determinado por un puntaje Z= ±3,18, para un valor de α =0.05 (en una prueba de dos extremos). A medida que aumentan los *gl*, las diferencias entre la distribución *t* y la curva normal se hacen cada vez menores. Cuando la muestra contiene más de 30 observaciones, la distribución *t* tiene idénticas propiedades que la distribución normal, ya que, a partir de este punto, las diferencias en las áreas de rechazo de la H_0 no tienen importancia práctica.

Al calcular un valor *t*, este debe ser comparado con los valores normativos tabulados para determinar si se rechaza o no la H_0. La correspondiente tabla contiene información sobre los grados de libertad (columnas) y a los niveles de significación (filas), junto con la información si la prueba es de uno o dos extremos. Entonces, dado un valor determinado de α y los grados de libertad, se compara el resultado del valor *t* obtenido con el valor tabulado. Si el valor *t* calculado es igual o mayor que el registro en la tabla, puede rechazarse la H_0 al correspondiente nivel de significación indicado.

Cualquier paquete estadístico calcula las probabilidades asociadas a un valor *t* dado, y sus correspondientes *gl* bajo el supuesto de que la H_0 es verdadera, de modo que no es necesario consultar los valores de tabla para tomar la decisión de rechazar o no esa hipótesis. Más adelante, en el desarrollo del ejemplo, veremos con más detalle este concepto.

Resultados de un experimento y prueba *t*

La distribución t puede tomarse como un modelo para describir la distribución de los resultados posibles en un experimento (suponiendo cierta la hipótesis nula), en varias pruebas de significación estadística. Por ejemplo, puede realizarse una prueba de hipótesis sobre el valor de una media paramétrica con muestras pequeñas. En este caso se calcula el valor *t* correspondiente a la media observada en una muestra aleatoria de la población a la que se refiere la hipótesis. La fórmula para el cálculo es la siguiente:

$$t_o = \frac{(\bar{x}-\mu)}{\delta_{\bar{x}}}$$

donde
μ es la media de la población bajo el supuesto que la H_0 es verdadera,
$\bar{x}$ es la media de una muestra aleatoria extraída de esa población, bajo el supuesto que la H_0 es verdadera,
$\delta_{\bar{x}}$ es el error estándar de la media, y se calcula mediante la siguiente ecuación:

$$\delta_{\bar{x}} = \frac{s}{\sqrt{N-1}}$$

Para determinar si el resultado de la media muestral se encuentra comprendido en la zona de aceptación o rechazo de la H_0, el valor *t* se referirá a una distribución t con *gl*=N-1.

Nótese que se está describiendo la situación en la que los valores de la variable en estudio son conocidos para una población, éstos son los que se comparan con los valores observados en una muestra aleatoria extraída de esa población. Entonces, para realizar la prueba de hipótesis debe aceptarse que la variable en estudio se distribuye de manera normal, o por lo menos de manera unimodal y simétrica en la población. Si se sabe o se sospecha que la variable no sigue esa distribución en la población, convendrá contar con una muestra grande y utilizar la distribución normal para describir la distribución de muestreo.

Diferencia entre medias muestrales

La prueba *t* puede emplearse como prueba de hipótesis sobre diferencia entre medias con muestras pequeñas. La típica disposición experimental para poner a prueba hipótesis sobre diferencias entre medias requiere disponer de dos muestras. Como se mencionó al inicio una de ellas se denomina grupo experimental y es examinada en determinadas condiciones experimentales, cuyos efectos son el objeto de estudio. El otro grupo se denomina control, y es observado en las mismas condiciones que el grupo experimental, pero en ausencia del efecto experimental que se quiere comprobar. En nuestro ejemplo, se desea verificar las prestaciones de un método especial para la enseñanza de la ortografía. El experimento consistió en la asignación aleatoria de escolares a dos grupos, en donde el grupo experimental quedó conformado por todos los niños que recibieron instrucción con el método nuevo, y el grupo control, por niños que recibieron instrucción con el método tradicional.

En las condiciones descriptas, la H_0 sostendría que no existirían diferencias entre los escolares que aprenden con un método especial de enseñanza de la ortografía y aquellos que no son instruidos con ese método. Estadísticamente, esto se plantearía como sigue:

H_0: $\mu_1 = \mu_2$
o bien
$\mu_1 - \mu_2 = 0$

Los subíndices 1 y 2 indican los grupos experimental y control respectivamente. La hipótesis alternativa, la cual da sentido al experimento, indica que existirán diferencias entre los grupos. Ahora bien, tales diferencias pueden estar no predichas (hipótesis bidireccional), o bien puede preverse el sentido de la diferencia (hipótesis unidireccional). En el primer caso, el planteo estadístico es el siguiente:

H_1: $\mu_1 \neq \mu_2$

En el segundo caso el planteo será:

H_1: $\mu_1 > \mu_2$

En una prueba que implica diferencia entre medias, el problema que se plantea es decidir si las diferencias observadas son debidas a las variaciones aleatorias del conjunto de datos, o bien, que además de esas variaciones se introdujo una variación sistemática, producto de la manipulación experimental. En otras palabras, el problema consiste en decidir si la manipulación de la variable experimental (el tipo de método empleado para la enseñanza de la ortografía), logró introducir una variación sistemática en el conjunto de datos, y si esas variaciones son diferentes a las que pudieran haberse obtenido por azar. La H_0 estipula que, de existir diferencias entre las medias de ambos grupos, serán debidas a factores asistemáticos, por lo que estás, o bien se anularán entre sí, o no resultarán estadísticamente significativas. Otra forma de decir esto es: la diferencia que se registra entre los grupos, podría observarse si se extrajeran aleatoriamente dos muestras de una misma población (siempre bajo el supuesto de la H_0).

Se habla de dos o más muestras independientes, cuando son seleccionadas de forma aleatoria de dos o más poblaciones. En otro caso, cuando las muestras son tomadas de la misma población, pero asignadas al azar a dos tratamientos diferentes. El término tratamiento en este contexto, debe entenderse como el efecto de la variable experimental, que debe ser una condición especificada con anterioridad.

Las muestras serán dependientes o estarán correlacionadas si un mismo grupo es analizado dos o más veces en condiciones diferentes. En este tipo de investigaciones, cada miembro de la muestra se transforma en su propio control. El ejemplo más frecuente de investigaciones de esta naturaleza, es cuando los sujetos son analizados antes y después de un tratamiento (v.g. psicoterapia, drogas, entrenamiento, etc.). También se habla de muestras correlacionadas cuando se procede a formar parejas de sujetos, que están igualados en una serie de variables de importancia para el estudio. Luego estos sujetos son asignados aleatoriamente a un grupo de tratamiento diferente.

Cuando se trata de un mismo sujeto, evaluado antes y después de un tratamiento, las mediciones estarán correlacionadas. El grado de correlación entre estas dos series de mediciones se expresa por el coeficiente de asociación de Pearson, que aparece en la fórmula del error estándar de la diferencia entre medias para muestras correlacionadas. Al tratarse con muestras pareadas, se tendrá en cuenta el coeficiente de correlación entre los valores de cada uno de los miembros del par.

Cálculo del estadístico t para muestras independientes

En la prueba de diferencia entre medias, el estadístico *t* se calcula con la siguiente fórmula:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\delta_{dif}}$$

En esta fórmula μ_1 y μ_2 representan las medias de las poblaciones representadas por las muestras. Como la H_0 sostiene que ambas medias son iguales, la expresión (μ_1 - μ_2) resulta cero o muy próximo a ese valor; $\bar{x}_1$y $\bar{x}_2$, son las medias de las muestras observadas, de ser cierta la H_0 la diferencia entre ambas será muy poca considerando que ambas fueron extraídas de manera aleatoria de la misma población; δ_{dif} representa el error estándar de las diferencias entre las medias. Como se observa, la fórmula puede ser simplificada de la siguiente manera:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\delta_{dif}}$$

Tratándose de muestras independiente el error estándar de la diferencia se calcula con la siguiente fórmula:

$$\delta_{dif} = \sqrt{\delta_{\bar{x}_1}^2 + \delta_{\bar{x}_2}^2}$$

los grados de libertad son $gl = N_1 + N_2 - 2$

Supuestos

La realización de una comparación entre medias, debe tener en cuenta dos supuestos básicos:

a) que, en las poblaciones representadas por las muestras, la variable en estudio se distribuye de manera normal,

b) que, en esas poblaciones la variabilidad de las variables es la misma (homogeneidad de varianza)

El primer supuesto es generalmente difícil de verificar, no obstante, los resultados de la prueba *t* resultan válidos aun cuando los supuestos no se cumplan estrictamente. Sin embargo, es de tener en cuenta que las variables en estudio deben distribuirse de modo unimodal y aproximadamente simétrico en la población. El supuesto de igualdad de varianza, es más fácil de verificar por medio de la prueba F de Snedecor (test de Levene).

Cuando existen sospechas de que las condiciones reales se apartan de manera importante de las condiciones supuestas, no debería emplearse una prueba *t*, sino algún similar no paramétrico.

Cuando las muestras están correlacionadas, se puede realizar una prueba de hipótesis acerca de la significación del coeficiente de correlación de Pearson. En estos casos, la hipótesis nula supone que: H_0: ρ=0, esto es, en la población de donde proviene la muestra, la correlación entre las variables X e Y es cero. La hipótesis de investigación H_1 en este caso supone que H_0: ρ≠ 0 esto es, que la correlación entre las variables X e Y es distinta de cero. Tal como está planteado, se trata de una hipótesis bidireccional. Cuando la hipótesis es unidireccional, supone una correlación positiva ente las variables (H_1= ρ> 0), o bien una correlación negativa (H_1= ρ< 0).

El estadístico *t* en este caso, se calcula con la fórmula:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

los grados de libertad son *gl*= N – 1

Este mismo cociente se puede emplear cuando se tiene un coeficiente de correlación de Spearman, con tal que N>10.

Esta prueba de significación supone que las variables correlacionadas se distribuyen de manera normal en la población. Como en el caso anterior, si se sospechan desviaciones importantes de este supuesto, convendrá abstenerse de efectuarla.

Cuando se trata de muestras correlacionadas, el cálculo del erro estándar de la diferencia entre las medias incluye el coeficiente de correlación de Pearson. La fórmula en este caso es la siguiente:

$$\delta_{dif} = \sqrt{\delta_{\bar{x}_1}^2 + \delta_{\bar{x}_2}^2 - 2r_{1,2}\delta_{\bar{x}_1}\delta_{\bar{x}_2}}$$

$\delta_{\bar{x}_1}\delta_{\bar{x}_2}$representan el error estándar de las medias de las muestras 1 y 2 respectivamente. Los grados de libertad en este caso se calculan con: *gl*= N – 1.

Resolución del problema

Según lo planteado al inicio, en una escuela se ha enseñado ortografía con el mismo método durante varios años y algunos docentes han elaborado un nuevo método que confían es mejor. Deciden realizar un experimento para comprobar esta suposición. Diseñan un experimento y seleccionan dos muestras aleatorias de alumnos: el grupo experimental, quedó conformado por todos los niños que recibieron instrucciones con el nuevo método de aprendizaje de la ortografía. El grupo control, quedó conformado por los niños que fueron instruidos con el método tradicional. Las restantes condiciones del experimento se mantuvieron exactamente iguales para ambos grupos. Es decir, la única diferencia entre ambos grupos es el método de instrucción con el que aprenden la ortografía.

Bajo estas circunstancias, la diferencia entre grupos puede expresarse bajo las siguientes hipótesis estadísticas:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

De ser verdadera la hipótesis nula se espera que no existan diferencias en el rendimiento promedio en las pruebas de ortografía del grupo control y el grupo experimental (o que, si existe alguna diferencia sea atribuida al azar). Para someter a prueba la H_0 los docentes compararon el rendimiento de los grupos mediante una prueba t de Student, para muestras independientes. Los resultados que se observan a continuación, fueron obtenidos mediante el uso del programa PSPP[1] y corresponde a la salida de una comparación de dos medias.

Modelo

```
T-TEST /VARIABLES= rendimiento
            /GROUPS=grupo(1,2)
            /CRITERIA=CI(0.95).
```

El recuadro detalla el modelo utilizado para el análisis: se trata de una prueba *t* para la variable rendimiento en una prueba de ortografía para dos grupos (1 y 2), prueba bidireccional con un nivel de confianza para ET 1= α =0.05 (Intervalo de Confianza=0.95)

Estadísticas de grupo: rendimiento en prueba de ortografía

grupo	**N**	**Media**	**Desviación Estándar**	**E.E.Media**
1=control	10	6,00	1,05	0,33
2=experimental	10	7,10	0,74	0,23

La siguiente tabla detalla los estadísticos obtenidos: tamaño de la muestra de cada grupo. La media del grupo control y la del grupo experimental. La desviación y el error estándar para cada grupo respectivamente. Lo que se deduce de esta tabla es que existe una diferencia en el rendimiento promedio de los grupos, aunque no es muy pronunciada.

Resultado de la prueba *t*

Prueba de Levene para la igualdad de varianzas		**Prueba *t* para la Igualdad de Medias**				
F	Sign.	t	df	Sign. (2-ext)	Diferencia Media	E.E. Diferencia
1,09	0,31	-2,70	18,00	0,015	-1,10	0,41

[1] Para obtener el programa PSPP de manera gratuita deben dirigirse a la siguiente dirección web: https://www.gnu.org/software/pspp/

En esta tabla se aprecia que la prueba de Levene, permite asumir homogeneidad de varianzas, uno de los supuestos principales para interpretar la diferencia entre medias encontrada. El valor *t* observado corresponde a una distribución con *gl*=18. La probabilidad de obtener una diferencia de medias de 1,10 puntos en las pruebas de rendimiento en ortografía, suponiendo que la H_0 es verdadera es igual a p=0,015. Esta probabilidad es muy baja; otra manera de interpretar este valor es por su inverso; la probabilidad de que la H_0 no sea verdadera es p=1-0,015= 0,985.

En conclusión, el método elaborado por los docentes produce una mejora en el aprendizaje de la ortografía de los escolares, aunque la magnitud de la diferencia encontrada es, en promedio de poco más de un punto.

Análisis de la varianza ANOVA

La diferencia encontrada entre los grupos según el método tradicional y el nuevo, si bien es significativa, no representa una ganancia considerable de uno respecto al otro. Los docentes decidieron emplear un nuevo método, que toma parte de los dos métodos antes analizados. Ahora se quiere conocer la eficacia de tres métodos diferentes para el aprendizaje de la lectura; llámense a estos métodos Tradicional (1), Nuevo (2) y Mixto (3). Siendo los métodos las variables independientes, se desea conocer cuál o cuáles de ellos resulta más eficaz para que los niños aprendan ortografía. Los puntajes de una prueba serán la variable dependiente.

El procedimiento para evaluar la eficiencia de los métodos requiere, como en el caso anterior, asignar aleatoriamente una cantidad N de niños a cada uno de los métodos en estudio, y luego comparar si existen diferencias entre ellos en el rendimiento en la prueba. Este procedimiento se resume a una comparación entre medias. El ANOVA permite comparar simultáneamente las tres medias (una para cada método), y determinar entre cuáles de ellas existe una diferencia significativa.

Precisiones terminológicas

Los escolares seleccionados para este experimento serán asignados aleatoriamente a un grupo (1, 2, 3), que recibirá instrucción con un método en particular. Cada método se denomina *tratamiento* y estos definen los niveles de la variable independiente. Por lo tanto, cada una de las categorías se denominan *niveles del factor*. Para este ejemplo, el tratamiento es el método utilizado para la enseñanza de la ortografía y sus niveles de factor serán Tradicional, Nuevo y Mixto. La técnica estadística pretende revelar si existen diferencias en los resultados obtenidos en la prueba de ortografía debida los tratamientos, o bien, las diferencias encontradas se deben al azar. Por la experiencia previa se sabe que parte de la variabilidad en los datos se debe al tipo de método utilizado, pero en este nuevo experimento se ha sumado un nuevo método que combina parte de los dos anteriores. Por lo tanto, se necesita comparar simultáneamente el rendimiento de tres grupos. El objetivo del análisis de la varianza es separar la varianza total, en dos componentes aditivos: a) la debida a los efectos de la variable experimental (independiente), y b) la debida a factores que no son considerados por el modelo. Para comprender esto vamos a introducir la terminología estadística apropiada:

X*ij*: X es el valor que asume la variable dependiente, el subíndice *i* designa a un individuo dado y el subíndice *j* designa al tratamiento que ha sido asignado. Así $\mathbf{X}_{3,1}$, designa el valor de la variable dependiente del tercer sujeto asignado al tratamiento 1.

ξ_j: representa el promedio del grupo j. Así, ξ_1 representará el promedio de los sujetos asignados al tratamiento 1, en la prueba de ortografía.

ξ: es el promedio general de todos los individuos independientemente del grupo al que haya sido asignado.

Para un individuo dado, su diferencia respecto al promedio general dependerá del tratamiento al que fue asignado, más los factores aleatorios que influyeron en el momento de la evaluación u otros que pudieran haber ejercido alguna influencia. Esta diferencia se representa por: la diferencia entre el puntaje obtenido por el individuo *i* en el tratamiento *j*, más la diferencia entre el promedio de todos los individuos asignados al tratamiento *j* y el promedio general de todos los tratamientos. Matemáticamente se representa por:

$$X_{ij} - \xi = \left(X_{ij} - \xi_j\right) + \left(\xi_j - \xi\right)$$

El primer miembro es la diferencia entre un valor individual y el promedio general de todos los grupos, por tanto, mide cuánto se aparta un individuo de la media general. En el segundo miembro, se representa cuánto se aparta un individuo del promedio de su grupo y estima las diferencias individuales. El tercer miembro representa cuánto que se aparta el grupo (en su promedio) del promedio general, y esta diferencia es atribuible al tratamiento aplicado.

La operación puede hacerse extensiva a todos los individuos de la muestra, pero por tratarse de desvíos en torno a la media, el resultado será cero. Por ello, es necesario considerar el cuadrado de estas diferencias. La fórmula en este caso se expresa por:

$$\xi\left(X_{ij} - \xi\right)^2 = \xi\left(\xi_j - \xi\right)^2 nj + \xi\left(X_{ij} - \xi_j\right)^2$$

Descomponiendo la fórmula tenemos que:

$\xi\,(X_{ij} - \xi)^2$ se denomina suma de cuadrado total (SCT).
$\xi\,(\xi_j - \xi)^2$ se denomina suma de cuadrados explicada (SCExp.) por el tratamiento, también llamada varianza intergrupal.
$\xi\,(X_{ij} - \xi_j)^2$ se denomina suma de cuadrado residual (SCRes.), también llamada varianza intragrupal.

La expresión *nj* en la suma de cuadrados explicada cumple la función de tener en cuenta los diferentes tamaños de los tratamientos, asignándoles un peso diferencial a cada uno de ellos; se conoce como *factor de ponderación*. Ahora se puede asignar a cada miembro su respectiva denominación:

Suma de Cuadrados Total: $\xi\,(X_{ij} - \xi)^2$
Suma de Cuadrados Explicada: $\xi\,(\xi_j - \xi)^2\, x\ nj$
Suma de Cuadrados Residual $\xi\,(X_{ij} - \xi_j)^2$

Así la expresión puede resumirse como: SCT = SCExp + SCRes

Estas expresiones se refieren respectivamente: a) variación total, b) variaciones que pueden atribuirse al modelo que se pone a prueba (eficacia de los tratamientos o métodos de aprendizaje), y c) variaciones debidas a diferencias individuales que son residuales y que por tanto no están explicadas por el modelo.

El análisis de varianza compara la varianza explicada por el modelo, la residual y la total. Pero las sumas de cuadrados en sí misma, no constituyen varianza, son solo los numeradores de las mismas. Para obtener las varianzas será necesario dividir las sumas de cuadrados por sus grados de libertad.

Los grados de libertad

El tamaño total de la muestra se denota como *n*, la cual está dividida en *k* tratamientos o grupos, que pueden contener la misma cantidad de individuos o no. De esta manera, la suma de los individuos que componen los *k* tratamientos, será igual al tamaño de la muestra.

Para obtener la varianza es necesario calcular los grados de libertad para cada una de las sumas de cuadrados señalada. Estos grados de libertad, dependerán del número de casos en cada una de las sumas de cuadrados y del número de parámetros que deban estimarse para su cálculo. Así, para la suma de cuadrados total SCT, se cuentan los *n* valores de *x*, cuando solo se ha estimado la media general ξ. Los grados de libertad en este caso serán los *n* casos de la muestra, menos uno, que es el número de parámetros estimado. Así, los *gl* (SCT) = n–1.

En el caso de la suma de cuadrados residual tenemos *n* valores de *x* y *k* medias de grupos o tratamientos, que son cada una de las X_{ij}, que se han estimado. Por lo tanto, los grados de libertad serán n – *k*. Así *gl* (SCRes.)=n–*k*.

Por último, los grados de libertad para la suma de cuadrados explicada, contiene *k* variables (las medias de los tratamientos) y solo un parámetro que se estima (la media poblacional), por lo que los grados de libertad serán n–1. Así, *gl* (SCExp.)= n–1.

Descomposición de la suma de cuadrados total

El análisis de varianza se basa en la comparación de la dispersión encontrada entre los grupos que pueden atribuirse a los tratamientos y aquella que se origina dentro de cada grupo debidas a factores que no han sido considerados en el modelo (variación aleatoria).

Así, se tendrán como estimadores de las respectivas varianzas a las sumas de cuadrados divididas por sus correspondientes grados de libertad. Estos estimadores se conocen como Cuadrados Medios (CM), y asumirán los siguientes valores:

CMExp. = SCExp / k- 1 representa el estimador de la varianza debida a los *k* tratamientos,

CMRes. = SCRes. / n – k representa el estimador de varianza debida a factores ajenos al tratamiento y que representan diferencias individuales.

La hipótesis nula

Retomando el ejemplo dado al principio, interesa saber si los diferentes métodos de enseñanza de la ortografía (tratamientos), verdaderamente producen efectos diferentes. En otros términos, se pretende saber si las medias de las distintas muestras son diferentes porque fueron sometidas a un tratamiento en particular, o si estas diferencias son producidas por mero azar.

La H_0 se postulará en función de la diferencia entre medias de los distintos tratamientos, y por ello estipulará que de haber diferencia esta será mínima y negligible. Estadísticamente esto se representa por:

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$

Al contrario, la hipótesis alternativa planteará que al menos una de esas medias será diferente.

La H_0 podrá rechazarse si se comprueba que las diferencias debidas a los tratamientos (expresada por CMExp.), sea substancialmente más importante que las debidas al azar (expresada por CMRes.). Para decidir si el CMExp. es significativamente mayor que el CMRes. se toma su cociente:

CMExp. / CMRes

y se usa una distribución de probabilidad especial que se conoce como distribución F de Fischer. Esta distribución provee para cada nivel de significación dado y los grados de libertar del numerado y el denominador, un puntaje crítico que determina el valor a partir del cual, las diferencias han de considerarse significativas. En los programas estadísticos se ofrece el valor de probabilidad exacto asignado al valor F bajo la H_0.

Tabla de ANOVA

Esta tabla resume el procedimiento para su cálculo.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados Medios	Valor F observado
Explicada	**SCExp.**	***k* – 1**	**CMExp.**	**F = CMExp. / CMRes.**
Residual	**SCRes.**	**n – *k***	**CMRes.**	
Total	**SCT**	**n – 1**		

Primero se procede a calcular la suma de cuadrados, luego se dividen por los grados de libertad para obtener los cuadrados medios, luego se obtiene el cociente entre los cuadrados medios y se determina si el valor obtenido para ese cociente superó el valor crítico de F en la distribución de probabilidad. Si efectivamente éste es mayor, se rechaza la H_0, de lo contrario no se la rechaza.

Resolución del problema

Habíamos comprobado que la diferencia encontrada entre los grupos según el método tradicional y el nuevo, no representa una ganancia considerable de uno respecto al otro, por ello los docentes decidieron emplear un nuevo método, combinando las propiedades de ambos. Se asignaron aleatoriamente estudiantes a cada uno de los métodos, conformando tres grupos diferentes, según el método utilizado para la enseñanza de la ortografía: Tradicional (1), Nuevo (2) y Mixto (3). Utilizando el ANOVA es posible identificar cuál o cuáles de ellos resulta más eficaz para que los niños aprendan ortografía.

Por lo tanto, los puntajes de una prueba serán la variable dependiente. El resultado de este experimento se expone en las siguientes tablas:

Modelo

```
ANOVA /VARIABLES= rendimiento
/GROUPS=grupo(1,2,3)
/CRITERIA=CI(0.95).
/ONEWAY /VARIABLES= rendimiento BY grupo
```

Estadísticos Descriptivos: dependiente rendimiento en prueba de ortografía

					Intervalo de Confianza 95% para la Media			
Grupos	N	Media	D. E.	E. E.	Límite Inferior	Límite Superior	Mínimo	Máximo
1 (T)	10	6,00	1,05	0,33	5,25	6,75	5,00	8,00
2 (N)	10	7,10	0,74	0,23	6,57	7,63	6,00	8,00
3 (M)	12	8,58	1,08	0,31	7,89	9,27	7,00	10,00
Total	32	7,31	1,45	0,26	6,79	7,83	5,00	10,00

En la tabla se observa que el rendimiento de los tres grupos es diferente respecto a la media. Se mantiene la diferencia de 1.1 puntos entre el método tradicional y el método nuevo, pero ahora se observa que el método que combina ambos (al que llamamos mixto) resulta superior en 1.48 puntos respecto al método nuevo, y en 2,58 puntos respecto al tradicional. Es decir, se aprecia una diferencia mayor que la obtenida en la comparación del método tradicional y nuevo. Entre los estadísticos descriptivos aparece la desviación estándar (D.E.), que como se observa, no muestra variaciones sustantivas entre los grupos.

El error estándar y los límites del intervalo para la media, son estadísticos útiles para la estimación de parámetros, cuestión que no nos ocupa en este ejemplo. De mayor importancia resultan los valores mínimo y máximo alcanzado en cada grupo. La inspección de las columnas muestra la diferencia lograda por la aplicación del método mixto.

Aun cuando estas variaciones nos hacen suponer que el método mixto es mejor, debemos rechazar la hipótesis nula que estipula que tales variaciones son producto del azar. Dicho en términos más sencillos, podríamos estar tentados de atribuir las diferencias observadas al factor método de enseñanza de la ortografía, cuando en realidad se trata de una variación espuria. El resultado del análisis de la varianza servirá para determinar si es posible rechazar la hipótesis nula.

Prueba de Homogeneidad de Varianzas

Estadístico de Levene	gl1	gl2	Sign.
1,37	2	29	0,270

Como en el caso de la prueba *t* de Student, el estadístico de Levene nos informa sobre el cumplimiento del supuesto de igualdad de varianzas para los grupos analizados. Tal como se observa, dicho supuesto se cumple.

Resultados del ANOVA

	Suma de Cuadrados	df	Cuadrado medio	F	Sign.
Entre Grupos	37,06	2	18,53	19,32	0,000
Intra Grupos	27,82	29	0,96		
Total	64,88	31			

En la tabla se resumen los resultados del análisis de varianza. Los conceptos que hemos repasado anteriormente se encuentran explicitados en cada una de las columnas. El más importante para nuestro planteo del problema es el nivel de significación de la prueba F. Nos indica la probabilidad de haber obtenido un valor de F=19,32 bajo el supuesto de que la hipótesis nula es verdadera. Como se observa, dicha probabilidad es menor al 1/1000. Se desprende de esto que resulta inverosímil considerar que la diferencia encontrada entre los grupos se deba meramente a variaciones aleatorias, en tal caso, aceptamos que el tratamiento ha tenido un efecto sistemático sobre el rendimiento de los escolares. En otros términos, es plausible aceptar que el método mixto para la enseñanza de la ortografía resulta ser el mejor de los tres empleados.