Inferencia

Estadística

(Teoría y problemas)

I. Espejo Miranda
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
A. M. Rodríguez Chía
A. Sánchez Navas
C. Valero Franco

Edita: Servicio de Publicaciones de la Universidad de Cádiz
c/ Doctor Marañón, 3. 11002 Cádiz (España)
www.uca.es/publicaciones

ISBN: 978-84-9828-131-6

Inferencia Estadística (Revisión: Marzo 2007)
I. Espejo Miranda, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Márquez, A. M. Rodríguez Chía, A. Sánchez Navas, C. Valero Franco

http://www.uca.es/teloydisren

Capítulo 6

Análisis de la Varianza

1. Introducción

Este capítulo está dedicado al estudio de una de las herramientas más valiosas de la Inferencia Estadística, el Análisis de la Varianza (ANOVA). Desarrollada por Fisher en la década de 1920 para resolver diversos problemas agrícolas, ha llegado a instaurarse como una de las técnicas más utilizadas en el análisis estadístico de los diseños de experimentos aplicados a diferentes campos de las Ciencias Experimentales, Ciencias de la Salud o Ciencias de la Educación.

Las siguientes situaciones reales ilustran algunos problemas que pueden ser abordados a través del análisis de la varianza.

Ejemplo 6.1

- *Un agricultor está interesado en estudiar si la aplicación de un abono en distintas dosis influye en el rendimiento de su plantación de maíz.*

- *Un laboratorio investiga si existen diferencias en la efectividad que tienen varios fármacos para combatir una determinada enfermedad.*

- *El director de un colegio está interesado en comparar el rendimiento de tres métodos de aprendizaje sobre la comprensión lectora de sus alumnos.*
- *Un laboratorio desea mejorar el rendimiento de un proceso químico y para ello quiere averiguar los efectos que la temperatura y el tiempo de reacción ejercen sobre el mismo.*

En todos los ejemplos anteriores se pone de manifiesto la necesidad de estudiar la influencia de un determinado *factor* o grupo de factores sobre una variable respuesta. Para ello, se comparan los efectos que las distintas dosis o *niveles* del factor producen en dicha variable. Esos niveles del factor, también denominados *tratamientos*, pueden ser de naturaleza cuantitativa o cualitativa.

Ejemplo 6.2 *En el ejemplo del colegio, la variable bajo estudio o variable respuesta es el rendimiento en la comprensión lectora, y el factor que se está considerando es el método para la enseñanza de la lectura, con tres niveles correspondientes a cada uno de los métodos.*

En el caso del laboratorio que desea mejorar el rendimiento de un proceso químico, la variable respuesta es dicho rendimiento y los factores que se están estudiando son la temperatura y el tiempo de reacción.

Puede ocurrir que el efecto producido por un nivel de un determinado factor esté influenciado por los niveles de otros factores. En ese caso, se dice que existe una interacción entre dichos factores, es decir, no actúan independientemente.

Ejemplo 6.3 *El efecto producido por un determinado tiempo de reacción sobre el rendimiento de un proceso químico está influenciado por los distintos niveles de temperatura, y recíprocamente.*

Formalmente, el *Análisis de la Varianza* permite analizar y comparar experimentos en los que se describe una variable respuesta cuantitativa en función de otras variables, cuantitativas o cualitativas, denominadas *factores*. Según se esté estudiando un único factor o varios, se distingue entre ANOVA de un sólo factor o unifactorial, y ANOVA multifactorial. Este capítulo se centra en el estudio del ANOVA unifactorial.

2. Modelo ANOVA unifactorial

Se sabe que sobre una determinada variable, X, influye un factor, δ, con k niveles, cuyos efectos interesa estudiar. El problema que se plantea es analizar si los distintos niveles del factor bajo estudio influyen de igual modo en la variable respuesta.

Para comprender el fundamento del análisis de la varianza en el caso unifactorial, se propone el siguiente ejemplo. A través de él se introducen, de modo intuitivo, las claves del modelo.

Ejemplo 6.4 *Se está realizando un estudio sobre el desarrollo y crecimiento de las doradas que se crían en las distintas piscinas de una piscifactoría.*

Se sabe que en el crecimiento de las doradas influyen diversas causas o factores, como la salinidad, la oxigenación, la temperatura del agua, la luz, el tipo de alimentación, así como la cantidad diaria de la misma, la propia constitución genética del pez, etc. Todos ellos actúan de un modo u otro, con mayor o menor intensidad, provocando la existencia de diferencias en el crecimiento.

En concreto, el estudio se centra en los efectos que, sobre el crecimiento de las doradas, ejerce una proteína que se les suministra diariamente como complemento a su alimentación. Si ésta es suministrada en distintas dosis, ¿los efectos de las dosis sobre el crecimiento serán diferentes?

Se trata de un problema típico del análisis de la va-

rianza donde sólo se estudia un factor, la proteína, y los efectos que producen distintos niveles de ésta en el crecimiento de las doradas.

Evidentemente, si la aplicación de los diferentes niveles de esta proteína producen un efecto análogo en el crecimiento de las doradas, se deducirá que dicha proteína no influye en el crecimiento. Por tanto, el objetivo será contrastar la hipótesis de que todos los niveles del factor actúan de igual forma.

Al estudiar la variabilidad que aparece en el crecimiento de las doradas sometidas a distintos niveles de proteína, debe considerarse que ésta puede deberse tanto al nivel de la proteína como a cualquiera de los restantes factores. Por tanto, a la hora de plantear el problema habrá que tener este hecho en cuenta.

Para abordar este tipo de problemas, el modelo que se propone es el siguiente: puesto que la variabilidad de la variable respuesta se debe a multitud de factores, se trata de dividir esa variabilidad en dos partes, la originada por el factor bajo estudio y la variación producida por el resto de factores, conocidos o no, denominada *error experimental*. Teóricamente esto es posible, como se verá más adelante.

En definitiva, el ANOVA nos va a permitir decidir si los distintos niveles del factor establecen diferentes subpoblaciones en la variable respuesta, o, por el contrario, el comportamiento de la variable respuesta es la misma para todos los niveles y se tiene una única población.

Como ha quedado de manifiesto, el problema de averiguar si los distintos niveles del factor afectan de igual modo a la variable respuesta, se plantea como un contraste de hipótesis. Para lo cual, habrá que utilizar algunas de las técnicas ya vistas en los capítulos anteriores.

El primer paso será expresar matemáticamente el modelo que se está estudiando, para, a continuación, descomponer la variabilidad total de las observaciones y realizar el correspondiente contraste de hipótesis.

2.1. Especificación del modelo

Se desea comparar los efectos de k tratamientos o niveles de un único factor sobre una determinada variable.

En este modelo se distinguen la variable cuantitativa respuesta, X, y la variable independiente o factor, δ. Supóngase que dicho factor se corresponde con una variable cualitativa con k niveles, $(\delta_1, \ldots, \delta_k)$.

De cada nivel i, se extrae una m.a.s. de tamaño n_i. El tamaño de cada una de las muestras no tiene por qué ser el mismo, aunque cuando sí lo es, el modelo se dice *equilibrado* y tiene ventajas sobre los *no equilibrados*, como se comentará más adelante.

La siguiente tabla recoge los $N = \sum_{i=1}^{k} n_i$ valores obtenidos en las k muestras de tamaños $n_1, \ldots, n_k$.

Nivel	Observaciones	Total	Media muestral	Media poblacional
1	$x_{11}\ x_{12}\ \ldots\ x_{1n_1}$	$x_{1.}$	$\overline{x}_1$	μ_1
2	$x_{21}\ x_{22}\ \ldots\ x_{2n_2}$	$x_{2.}$	$\overline{x}_2$	μ_2
$\vdots$	$\vdots\ \vdots\ \ \vdots$	$\vdots$	$\vdots$	$\vdots$
k	$x_{k1}\ x_{k2}\ \ldots\ x_{kn_k}$	$x_{k.}$	$\overline{x}_k$	μ_k
Valores globales		$x_{..}$	$\overline{x}_{..}$	μ

Donde $x_{i.}$ representa el total de las observaciones en el i-ésimo nivel, es decir, $\sum_{j=1}^{n_i} x_{ij} = x_{i.}$, y $\overline{x}_{i.}$ es el promedio de las observaciones en dicho nivel i, es decir, $\overline{x}_{i.} = \frac{x_{i.}}{n_i}$.

A su vez, la suma de todas las observaciones viene dada por

$$x_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij},$$

y la media general de todas ellas por

$$\overline{x}_{..} = \frac{x_{..}}{N}.$$

El modelo de análisis de la varianza de un sólo factor se expresa:

$$X_i = \mu_i + \varepsilon_i \qquad i = 1, \ldots, k$$

donde μ_i sería la parte de X_i explicada por el factor, la media de la población del nivel i, y ε_i es la variación causada por los factores desconocidos o no estudiados. Este modelo también se puede expresar como:

$$X_i \quad = \quad \mu + \delta_i + \varepsilon_i, \tag{6.1}$$

donde $\mu_i = \mu + \delta_i$, siendo:

μ: la media global de la población, parámetro común a todos los niveles.

δ_i: el efecto diferencial respecto a la media global del i-ésimo nivel del factor sobre la respuesta.

ε_i: la componente aleatoria del error, la parte de la variable no explicada por μ y δ_i, es decir, el efecto de los factores extraños o no controlables.

El modelo definido por (6.1), describe dos situaciones con respecto al efecto de los tratamientos. Cuando los k tratamientos son seleccionados específicamente por el experimentador, se dice que el modelo es de *efectos fijos* y sus conclusiones no pueden hacerse extensibles a otros niveles similares que no hayan sido considerados específicamente. Por el contrario, cuando los k tratamientos constituyen una muestra aleatoria de una población de tratamientos, se dice que el modelo es de efectos aleatorios o *modelo aleatorio*. En este último caso, el objetivo será generalizar las conclusiones extraídas de una muestra a todos los tratamientos de la población. Cuando hay más de un factor, puede ocurrir que unos sean fijos y otros aleatorios, en ese caso se denomina *modelo mixto*.

Nuestro estudio se centra en el modelo de efectos fijos. En dicho modelo, puesto que los k niveles del factor son todos los que existen, los efectos δ_i son constantes, es decir, no son variables aleatorias. De esta forma, en el modelo

$$X_i = \mu + \delta_i + \varepsilon_i,$$

las únicas variables aleatorias son los X_i y los ε_i.

Por otra parte, deben verificarse las siguientes hipótesis:

- Los ε_i son variables aleatorias independientes.
- Los ε_i se distribuyen normalmente.
- $\mathrm{V}[\varepsilon_i] = \sigma^2$, siendo σ constante para todos los niveles del factor.

Estas hipótesis son fundamentales para el buen funcionamiento del modelo, por lo que antes de aplicarse el ANOVA deberá comprobarse que se verifican. Por esta razón, se ha dedicado un apartado a analizar la validación de dichas hipótesis.

En estas condiciones se tiene:

1. $\mathrm{E}[X_i] = \mathrm{E}[\mu_i + \varepsilon_i] = \mu_i$
2. $\mathrm{V}[X_i] = \mathrm{V}[\mu + \delta_i + \varepsilon_i] = \sigma^2$
3. X_i son independientes ya que ε_i son independientes, y por tanto $X_i \sim N(\mu_i, \sigma)$.

Puesto que el estudio que se está realizando se centra en los efectos de los distintos niveles del factor sobre la variable respuesta, interesa probar hipótesis sobre los δ_i, $i = 1, \ldots, k$. Para lo cual, habrá que estimar previamente los parámetros del modelo.

2.2. Estimación de los parámetros

En el modelo ANOVA de efectos fijos, a partir de la única información disponible, es decir, las observaciones x_{ij}, se tienen que estimar μ, σ^2 y δ_i, $i = 1, \ldots, k$.

Intuitivamente, dado que $\delta_i = \mu_i - \mu$, si se estima μ y μ_i por las respectivas medias muestrales, esto es, $\overline{x}_{..}$ y $\overline{x}_{i.}$, se tiene que la estimación de δ_i será $\overline{x}_{i.} - \overline{x}_{..}$.

Otra forma de obtener estimadores de μ y δ_i es recurriendo al método de mínimos cuadrados, para ello:

$$L(\mu, \delta_1, \ldots, \delta_k) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \varepsilon_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \mu - \delta_i)^2$$

$$\left.\frac{\partial L}{\partial \mu}\right|_{\hat{\mu}, \hat{\delta}_i} = 0$$

$$\left.\frac{\partial L}{\partial \delta_i}\right|_{\hat{\mu}, \hat{\delta}_i} = 0; \qquad i = 1, \ldots, k.$$

Obteniéndose como ecuaciones normales:

$$\begin{array}{lclclclclcl}
N\hat{\mu} & + & n_1\hat{\delta}_1 & + & n_2\hat{\delta}_2 & + & \cdots & + & n_k\hat{\delta}_k & = & x_{..} \\
n_1\hat{\mu} & + & n_1\hat{\delta}_1 & & & & & & & = & x_{1.} \\
n_2\hat{\mu} & & & + & n_2\hat{\delta}_2 & & & & & = & x_{2.} \\
 & & & & & & \vdots & & & & \\
n_k\hat{\mu} & & & & & & & + & n_k\hat{\delta}_k & = & x_{k.}
\end{array}$$

Debido a que las ecuaciones no son independientes, la solución no es única. Considerando los efectos de los niveles como desviaciones de la media general, se tiene la restricción:

$$\sum_{i=1}^{k} n_i \hat{\delta}_i = 0,$$

lo que da como solución del sistema,

$$\begin{array}{l}
\hat{\mu} = \overline{x}_{..} \\
\hat{\delta}_i = \overline{x}_{i.} - \overline{x}_{..}, \qquad i = 1, \ldots, k,
\end{array}$$

coincidiendo con las estimaciones obtenidas de forma intuitiva.

Para conseguir un estimador del error, $\varepsilon_{ij} = x_{ij} - \mu - \delta_i$, basta sustituir las estimaciones obtenidas de μ y δ_i, resultando:

$$\hat{\varepsilon}_{ij} = x_{ij} - \overline{x}_{..} - (\overline{x}_{i.} - \overline{x}_{..}) = x_{ij} - \overline{x}_{i.}$$

Para obtener un estimador de σ^2, se recurre al error experimental, ε_{ij}:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij}^2}{N}$$

Los estimadores de los ε_{ij}, $\hat{\varepsilon}_{ij}$, no son independientes, ya que están sometidos a las restricciones que resultan al estimar los parámetros. Es decir,

$$\begin{aligned}
\hat{\mu}_i &= \overline{x}_{i.} = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \\
\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij} &= \sum_{j=1}^{n_i} x_{ij} - n_i\overline{x}_{i.} = 0, \quad i = 1,\ldots,k.
\end{aligned}$$

Se tienen k ecuaciones de restricción y n_i valores de $\hat{\varepsilon}_{ij}$ que determinar en cada una de ellas. Puesto que $\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij} = 0$ para el efecto i-ésimo, bastará con conocer $n_i - 1$ valores de $\hat{\varepsilon}_{ij}$, pues el otro quedará inmediatamente determinado. Luego el número de valores no determinados a priori será $\sum_{i=1}^{k}(n_i - 1) = N - k$.

Dado que

$$\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij}^2}{\sigma^2} = \frac{N\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{x_{ij} - \overline{x}_{i.}}{\sigma}\right)^2$$

y debido a que las observaciones x_{ij} son independientes, se verifica que

$$\sum_{j=1}^{n_i}\left(\frac{x_{ij} - \overline{x}_{i.}}{\sigma}\right)^2 \sim \chi^2_{n_i - 1},$$

luego

$$\frac{N\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-k};$$

por tanto,

$$\mathrm{E}\left[\frac{N\hat{\sigma}^2}{\sigma^2}\right] = N - k \Rightarrow \mathrm{E}[\hat{\sigma}^2] = \frac{N-k}{N}\sigma^2,$$

con lo que un estimador insesgado de σ^2 es

$$\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij}^2}{N-k}$$

2.3. Contraste de hipótesis

Como ya fue mencionado anteriormente, en el modelo ANOVA de un factor con k niveles, interesa saber si todos los niveles del factor producen el mismo efecto. Es decir, en el modelo

$$X_i = \mu + \delta_i + \varepsilon_i,$$

se quiere contrastar la hipótesis de que $\delta_i = 0$, para $i = 1, \ldots, k$. Como $\delta_i = \mu_i - \mu$, en realidad el contraste que se plantea es si $\mu_i = \mu$, para $i = 1, \ldots, k$.

Se trata, pues, de realizar el contraste

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu \\ H_1 : \mu_i \neq \mu, \text{ para algún } i \in \{1, \ldots, k\}. \end{array} \right.$$

Cabría preguntarse si este contraste puede realizarse mediante comparaciones dos a dos por las técnicas vistas en capítulos anteriores. El problema radica en que no se puede realizar contrastes sobre la diferencia de dos medias de forma aislada, a un nivel de confianza $1 - \alpha$, ya que no se sabe con seguridad cuál será el nivel de confianza para el conjunto de todos los contrastes. Esto podría provocar, por ejemplo, el rechazar la hipótesis de igualdad de medias siendo cierta con una alta probabilidad.

Ejemplo 6.5 *En el estudio de los efectos de los 5 niveles de un determinado factor sobre una variable concreta, se desea probar la igualdad de las 5 medias, usando comparación por pares. El número de pares posibles es 10, y si la probabilidad de no rechazar la Hipótesis Nula en cada prueba es $1 - \alpha = 0'95$, la probabilidad de no rechazar la Hipótesis Nula en las 10 pruebas es $(0'95)^{10} = 0'60$, si son independientes. Se observa así un incremento considerable de la probabilidad de error tipo I.*

El contraste F que se propone se basa en la descomposición de la variabilidad total de los datos para encontrar el estadístico de contraste.

2.3.1. Descomposición de la variabilidad de los datos

Intuitivamente, si se está estudiando los efectos que producen los k niveles del factor sobre el conjunto de datos, la variabilidad total de esos datos puede deberse a dos posibles causas: por un lado, la variabilidad dentro de cada nivel, y por otra, a la variabilidad entre niveles.

Formalmente, tomando la (i, j)-ésima observación,

$$x_{ij} = \mu + \delta_i + \varepsilon_{ij}$$

y sustituyendo μ, δ_i y ε_{ij} por sus estimadores, se tiene

$$x_{ij} = \overline{x}_{..} + (\overline{x}_{i.} - \overline{x}_{..}) + (x_{ij} - \overline{x}_{i.})$$

y

$$(x_{ij} - \overline{x}_{..}) = (\overline{x}_{i.} - \overline{x}_{..}) + (x_{ij} - \overline{x}_{i.}).$$

Tomando las N observaciones y elevando al cuadrado (suma de cuadrados corregida),

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(\overline{x}_{i.} - \overline{x}_{..}) + (x_{ij} - \overline{x}_{i.})]^2$$

o bien,

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{..})^2}_{SC_T} = \underbrace{\sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..})^2}_{SC_{Tr}} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2}_{SC_E}.$$

Tomando como medida de la variabilidad total de los datos la suma de cuadrados corregida, la ecuación anterior muestra que dicha variabilidad puede descomponerse en la suma de cuadrados de las diferencias entre la media de los niveles o tratamientos y la media global, y la suma de cuadrados de las diferencias entre las observaciones dentro del nivel y la media en dicho nivel. Es decir,

$$SC_T = SC_{Tr} + SC_E,$$

donde:

SC_{Tr}: Suma de cuadrados debida a los tratamientos (entre tratamientos), que mide la diferencia entre niveles. Si todos los niveles medios son iguales, $SC_{Tr} = 0$. Cuanto más difieran los niveles medios, mayor será SC_{Tr}.

SC_E: Suma de cuadrados debida al error (dentro de los tratamientos), que mide la variabilidad total debida al azar o no controlada. Cuanto menor sea la variación entre las observaciones de cada nivel del factor, menor será SC_E. Si $SC_E = 0$, para cada nivel todas las observaciones son iguales.

Para poder estudiar y comparar la variabilidad de forma que no dependa de las unidades de medida, se calcula el cociente de ambas. Cuando existan diferencias entre los distintos niveles, el cociente será grande, mientras que cuando no existan será pequeño. Si fuese cierta la Hipótesis Nula, esto es, si $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$, las $\overline{x}_{i.}$ tendrían que ser muy parecidas y por tanto, cercanas a $\overline{x}_{..}$, estando SC_{Tr} muy próxima de cero y siendo pequeña en relación con SC_E. Luego si H_0 es cierta, $\frac{SC_{Tr}}{SC_E}$ será pequeño. Por tanto, parece que este cociente o alguna transformación del mismo, podría ser un estadístico apropiado para el contraste.

2.3.2. Estadístico de contraste

A partir del modelo,

$$X_i = \mu + \delta_i + \varepsilon_i,$$

se tiene que

$$x_{i.} = \sum_{j=1}^{n_i} x_{ij} = n_i\mu + n_i\delta_i + \varepsilon_{i.}$$

$$\overline{x}_{i.} = \mu + \delta_i + \overline{\varepsilon}_{i.}$$

$$x_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} = N\mu + \sum_{i=1}^{k} n_i\delta_i + \varepsilon_{..}.$$

Como $\sum_{i=1}^{k} n_i\delta_i = 0$,

$$\overline{x}_{..} = \mu + \overline{\varepsilon}_{..}.$$

Luego,

$$x_{ij} - \overline{x}_{..} = \delta_i + (\varepsilon_{ij} - \overline{\varepsilon}_{..})$$

$$x_{ij} - \overline{x}_{i.} = \varepsilon_{ij} - \overline{\varepsilon}_{i.}$$

$$\overline{x}_{i.} - \overline{x}_{..} = \delta_i + (\overline{\varepsilon}_{i.} - \overline{\varepsilon}_{..}).$$

A partir de estas expresiones, cada una de las componentes de la variabilidad total se puede expresar del modo siguiente:

1. La variabilidad total de los datos, SC_T.

$$\begin{aligned} SC_T &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{..})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[\delta_i + (\varepsilon_{ij} - \overline{\varepsilon}_{..})]^2 \\ &= \sum_{i=1}^{k} n_i\delta_i^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\varepsilon_{ij} - \overline{\varepsilon}_{..})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n_i}\delta_i(\varepsilon_{ij} - \overline{\varepsilon}_{..}). \end{aligned}$$

Hay un total de N desviaciones $x_{ij} - \overline{x}_{..}$ que no son independientes, ya que

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{..}) = 0,$$

por lo que el número de grados de libertad de SC_T es $N-1$.

Si se calcula el valor esperado de $\frac{SC_T}{N-1}$, teniendo en cuenta que $\varepsilon_i \sim N(0,\sigma)$, se obtiene

$$\begin{aligned} \mathrm{E}\left[\frac{SC_T}{N-1}\right] &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{N-1} + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mathrm{E}[(\varepsilon_{ij} - \overline{\varepsilon}_{..})^2]}{N-1} \\ &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{N-1} + \frac{N}{N-1}\left(\frac{N-1}{N}\sigma^2\right) \\ &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{N-1} + \sigma^2. \end{aligned}$$

Denotando por MC_T a la media de cuadrados $\frac{SC_T}{N-1}$, se tiene que bajo H_0, es $\mathrm{E}[MC_T] = \sigma^2$. En ese caso, MC_T es insesgado para σ^2.

2. La medida de las diferencias entre niveles, SC_{Tr}.

$$\begin{aligned} SC_{Tr} &= \sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..})^2 \\ &= \sum_{i=1}^{k} n_i[\delta_i + (\overline{\varepsilon}_{i.} - \overline{\varepsilon}_{..})]^2 \\ &= \sum_{i=1}^{k} n_i\delta_i^2 + \sum_{i=1}^{k} n_i(\overline{\varepsilon}_{i.} - \overline{\varepsilon}_{..})^2 + 2\sum_{i=1}^{k} n_i\delta_i(\overline{\varepsilon}_{i.} - \overline{\varepsilon}_{..}). \end{aligned}$$

Aunque hay k niveles medios estimados, se está trabajando bajo el supuesto

$$\sum_{i=1}^{k} n_i\hat{\delta}_i = \sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..}) = 0,$$

luego SC_{Tr} tiene $k-1$ grados de libertad.

Si se calcula el valor esperado de $\frac{SC_{Tr}}{k-1}$,

$$\begin{aligned} \mathrm{E}\left[\frac{SC_{Tr}}{k-1}\right] &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{k-1} + \sum_{i=1}^{k} n_i\frac{\mathrm{E}[(\overline{\varepsilon}_{i.} - \overline{\varepsilon}_{..})^2]}{k-1} \\ &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{k-1} + \frac{1}{k-1}\sum_{i=1}^{k} n_i\left[\mathrm{E}[\overline{\varepsilon}_{i.}^2] + \mathrm{E}[\overline{\varepsilon}_{..}^2] - \mathrm{E}[\overline{\varepsilon}_{i.}\overline{\varepsilon}_{..}]\right] \\ &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{k-1} + \sum_{i=1}^{k} \frac{n_i}{k-1}\left[\frac{\sigma^2}{n_i} + \frac{\sigma^2}{N} - 2\frac{\sigma^2}{N}\right] \\ &= \sum_{i=1}^{k} n_i\frac{\delta_i^2}{k-1} + \sigma^2. \end{aligned}$$

Denotando $MC_{Tr} = \frac{SC_{Tr}}{k-1}$, se tiene que MC_{Tr} estima

$$\sum_{i=1}^{k} n_i\frac{\delta_i^2}{k-1} + \sigma^2.$$

Bajo H_0, es $\mathrm{E}[MC_{Tr}] = \sigma^2$, y en ese caso MC_{Tr} es estimador insesgado de σ^2.

3. La variabilidad no controlada, SC_E.

$$\begin{aligned} SC_E &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{i.})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\varepsilon_{ij}-\overline{\varepsilon}_{i.})^2. \end{aligned}$$

Puesto que dentro de cada nivel existen n_i observaciones, cada uno de ellos tiene asociado $n_i - 1$ grados de libertad. Como hay k niveles, el número de grados de libertad de SC_E es

$$(n_1 - 1) + (n_2 - 1) + \ldots + (n_k - 1) = N - k.$$

Si se calcula el valor esperado de $\frac{SC_E}{N-k}$, se tiene que

$$\begin{aligned} \mathrm{E}\left[\frac{SC_E}{N-k}\right] &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\mathrm{E}\left[(\varepsilon_{ij}-\overline{\varepsilon}_{i.})^2\right]}{N-k} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\frac{\sigma^2}{N-k}-\frac{1}{N-k}\frac{\sigma^2}{n_i}\right) \\ &= \sigma^2 \end{aligned}$$

Denotando $MC_E = \frac{SC_E}{N-k}$, se tiene que MC_E es estimador insesgado para σ^2.

Bajo las hipótesis del modelo ANOVA unifactorial que se está estudiando, es posible demostrar que:

- $\frac{SC_E}{\sigma^2}$ sigue una distribución Chi–cuadrado con $N-k$ grados de libertad.
- Si la Hipótesis Nula es cierta, $\frac{SC_{Tr}}{\sigma^2}$ sigue una distribución Chi–cuadrado con $k-1$ grados de libertad.

- SC_{Tr} y SC_E son independientes.

Por tanto, si la Hipótesis Nula es cierta, el estadístico

$$F_0 = \frac{\frac{SC_{Tr}}{\sigma^2}\frac{1}{k-1}}{\frac{SC_E}{\sigma^2}\frac{1}{N-k}} = \frac{MC_{Tr}}{MC_E}$$

sigue una distribución F de Snedecor con $k-1$ y $N-k$ grados de libertad.

Este estadístico proporciona, fijado un nivel de significación α, la siguiente regla de decisión del contraste:

$$\text{Rechazar } H_0 \text{ cuando } F_0 > F_{1-\alpha,k-1,N-k}.$$

Si se rechaza la Hipótesis Nula, esto es, que $\delta_i = 0$, para $i = 1, \ldots, k$, es porque al menos dos de los efectos son diferentes de cero, pudiendo ser cero el resto.

2.3.3. Tabla ANOVA

El proceso para realizar el contraste:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu \\ H_1 : \mu_i \neq \mu, \text{ para algún } i \in \{1, \ldots, k\} \end{cases}$$

se resume en la tabla 6.1, conocida como tabla ANOVA. Para facilitar los cálculos, se proponen las siguientes variantes de SC_{Tr} y SC_E:

$$SC_{Tr} = \sum_{i=1}^{k} n_i(\overline{x}_{i.} - \overline{x}_{..})^2 = \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}$$

$$SC_E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i}$$

Si $F_0 > F_{1-\alpha,k-1,N-k}$, se rechazará la Hipótesis Nula, es decir, que los efectos que producen los distintos niveles del factor sean iguales.

Fuente	Suma de cuadrados (SC)	Grados de libertad (G.L.)	Media de cuadrados (MC)	F_0
Entre (Tr)	$\sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}$	$k-1$	$\frac{SC_{Tr}}{k-1}$	
Dentro (E)	$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i}$	$N-k$	$\frac{SC_E}{N-k}$	$\frac{MC_{Tr}}{MC_E}$
Total	$SC_{Tr} + SC_E$	$N-1$		

Tabla 6.1: Tabla ANOVA

Ejemplo 6.6 *Un laboratorio investiga la efectividad de* 5 *nuevos medicamentos para combatir la hepatitis. Puesto que se dispone de una cantidad limitada de cada uno de ellos, sólo puede administrarse a* $4, 2, 2, 3$ *y* 1 *pacientes, respectivamente. Para medir la efectividad, se mide el tiempo de reacción después de inyectar cada medicamento. Para el experimento, se toma una muestra aleatoria de* 12 *personas y a cada uno se le inyecta una dosis de uno de los medicamentos, seleccionado aleatoriamente. Tras medir el tiempo de reacción los datos se recogieron en la siguiente tabla:*

Medicamento	*Tiempo de reacción*			
A	$8'3$	$7'6$	$8'4$	$8'3$
B	$7'4$	$7'1$		
C	$8'1$	$6'4$		
D	$7'9$	$9'5$	$10'0$	
E	$7'1$			

La variable respuesta es el tiempo de reacción, y se está estudiando un único factor, el medicamento, con 5 *niveles. El objetivo será averiguar si existen diferencias entre los efectos que producen los medicamentos. Así, si se identifican los respectivos medicamentos con los niveles* 1 *a* 5 *del factor, interesa realizar el siguiente contraste:*

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_5 \\ H_1: \mu_i \neq \mu_j, \text{ algún } i,j \in \{1,\ldots,5\}, \end{cases}$$

para el que se supondrá $\alpha = 0'05$. En primer lugar, hay que organizar los datos para su análisis.

Med.	*Tiempo reacción*	n_i	$x_{i.}$	$\sum_j x_{ij}^2$	$x_{i.}^2/n_i$
A	$8'3, 7'6, 8'4, 8'3$	4	$32'6$	$266'10$	$265'69$
B	$7'4, 7'1$	2	$14'5$	$105'17$	$105'13$
C	$8'1, 6'4$	2	$14'5$	$106'57$	$105'13$
D	$7'9, 9'5, 10'0$	3	$27'4$	$252'66$	$250'25$
E	$7'1$	1	$7'1$	$50'41$	$50'41$
Total		12	$96'1$	$780'91$	$776'61$

De donde

$$\begin{aligned} SC_{Tr} &= \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N} \\ &= 776'61 - \tfrac{96'1^2}{12} = 7'01 \end{aligned}$$

$$\begin{aligned} SC_E &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} \\ &= 780'91 - 776'61 = 4'30. \end{aligned}$$

Y la tabla ANOVA para este ejemplo vendría dada por:

Fuente	*SC*	*G.L.*	*MC*	F_0
Entre	$7'01$	4	$\frac{7'01}{4} = 1'75$	
Dentro	$4'30$	7	$\frac{4'30}{7} = 0'61$	$\frac{1'75}{0'61} = 2'84$
Total	$11'31$	11		

Como $F_{0'95,4,7} = 4'12 > 2'84 = F_0$, se concluye que no se puede rechazar la hipótesis de que los medicamentos tengan el mismo efecto.

2.3.4. Intervalo de confianza para la media

El contraste que se viene analizando permite determinar si difieren los niveles medios de los tratamientos, μ_i. Si dicho contraste conduce a que los niveles medios de los tratamientos son iguales, se concluye que desde la óptica de la variable respuesta la población es única.

Si la población es única, el objetivo que se plantea ahora es obtener un intervalo de confianza para la media de la misma. Para ello, en lugar de aislar las observaciones en una única muestra y utilizar la cuasivarianza muestral para estimar la varianza poblacional, se puede utilizar un estimador mejor mediante el MC_E, el cual refleja información de las muestras conjuntas e interviene en la varianza de la media muestral bajo cada tratamiento.

En concreto, se verifica que

$$\frac{\overline{x}_{i.} - \mu_i}{\sqrt{\mathrm{V}[\overline{x}_{i.}]}} \sim t_{N-k},$$

donde

$$\mathrm{V}[\overline{x}_{i.}] = \frac{\sigma^2}{n_i}.$$

Estimando σ^2 por MC_E, se tiene que, la estimación por intervalos de confianza, al nivel $1-\alpha$, para la media μ_i es:

$$IC_{1-\alpha}(\mu_i) = \left[\overline{x}_{i.} \pm t_{1-\frac{\alpha}{2},N-k}\sqrt{\frac{MC_E}{n_i}}\right].$$

Si en la realización del contraste no se tuvieran evidencias para rechazar la hipótesis de igualdad de medias de los diferentes tratamientos, se puede estimar la media común de todos los tratamientos a través de la media $\overline{x}_{..}$. Ya que esta verifica que

$$\frac{\overline{x}_{..} - \mu}{\sqrt{\mathrm{V}[\overline{x}_{..}]}} \sim t_{N-1},$$

y estimando de nuevo σ^2 por MC_E, se tiene que el intervalo de confianza, al nivel $1-\alpha$, para la media μ es:

$$IC_{1-\alpha}(\mu) = \left[\overline{x}_{..} \pm t_{1-\frac{\alpha}{2},N-1} \sqrt{\frac{MC_E}{N}} \right].$$

3. Contrastes múltiples

Cuando la realización del contraste:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu \\ H_1 : \mu_i \neq \mu, \text{ algún } i \in \{1, \ldots, k\}, \end{cases}$$

lleva a rechazar la Hipótesis Nula, es decir, a que los niveles medios de los tratamientos sean diferentes, se concluye que existe influencia entre el factor aplicado a la variable bajo estudio y la propia variable. Ahora bien, con el objetivo de realizar un estudio más detallado, se analizarán cuáles son los niveles que presentan medias diferentes.

Esto supone, bien estimar las medias individuales de cada una de las poblaciones, o bien comparar pares de medias de poblaciones en términos de sus diferencias estimadas.

Como ya se comentó anteriormente, los problemas surgen cuando se comparan pares de medias de forma aislada, ya que incrementan la probabilidad de error tipo I. Por esta razón, se proponen los denominados *procedimientos de comparaciones múltiples.*

Los procedimientos de comparaciones múltiples se pueden agrupar en dos categorías: procedimientos de comparaciones simultáneas , donde se utiliza el mismo estadístico con cada par de medias y procedimientos de comparaciones secuenciales, donde el estadístico depende del número de medias muestrales que se encuentran entre las dos que se contrastan, tras la ordenación previa de las k medias.

A continuación, el estudio se centra en el análisis de estos procedimientos con el fin de detectar diferencias significativas entre las medias de los distintos niveles del factor, cuando se ha rechazado la Hipótesis

Nula propuesta en el contraste del análisis de la varianza. De ahora en adelante se trabajará con un nivel de significación conjunto α.

3.1. Test de Duncan

Para efectuar comparaciones múltiples entre k medias, el test de Duncan requiere de la media de cuadrados del error, MC_E, sus grados de libertad y que el número de observaciones por nivel sea el mismo en todos ellos, es decir, $n_i = n_j = n$, $\forall i, j$.

Tras ordenar de forma creciente las k medias muestrales o promedios de tratamiento, se determina el error estándar en cada promedio usando

$$\hat{\sigma}_x = \sqrt{\frac{MC_E}{n}}.$$

La idea de este test es calcular las diferencias entre las dos medias a comparar, y averiguar si es mayor que una cierta cantidad que a continuación se calculará, R_p. Si es así, se rechazará que ambas medias sean iguales al nivel de probabilidad elegido; de las medias que están entre ellas no se puede decir nada, por lo que será necesario continuar el contraste. Si la diferencia en cambio es menor que R_p, no se puede rechazar que ambas medias sean iguales, sucediendo lo mismo con todas las contenidas entre ellas.

Para calcular R_p, siendo p el número de medias entre las dos a comparar, ambas inclusive, es necesario conocer el valor tabulado para el recorrido de Duncan, r_p, dado en la tabla A.44. Este viene dado por $r_p = C_{1-\alpha,p,N-k}$, siendo $N - k$ los grados de libertad del error. Una vez conocido r_p, el valor de R_p viene determinado por la expresión

$$R_p = r_p\sqrt{\frac{MC_E}{n}}.$$

Ejemplo 6.7 *Un grupo de psicólogos y naturalistas están estudiando las causas del insomnio para intentar encontrar el modo de solucionarlo. Tras muchas investigaciones, dieron con cinco métodos que facilitan el* *sueño: pastillas, hierbas naturales, autosugestión, yoga y gimnasia. Tras someter a varios pacientes que sufrían insomnio a cada uno de estos métodos, asignados aleatoriamente, las horas de sueño se reflejan en la siguiente tabla:*

Pastillas	*Yoga*	*Autosuges.*	*Hierbas*	*Gimnasia*
$9'34$	$6'46$	$5'79$	$8'37$	$4'94$
$8'53$	$4'83$	$5'13$	$7'57$	$4'11$
$9'43$	$5'89$	$6'17$	$8'69$	$5'45$
$8'37$	$5'30$	$4'72$	$8'06$	$5'21$
$9'64$	$6'33$	$5'60$	$7'23$	$5'00$

Se ha comprobado que existen diferencias entre los métodos. Se trata ahora de averiguar, con un nivel de significación de $0'05$*, cuáles son los métodos que presentan las diferencias.*

Se tienen $N = 25$ *observaciones y se desea comparar* $k = 5$ *medias. En primer lugar se calcularán dichas medias y se ordenarán de mayor a menor:*

$$\begin{array}{l} \overline{x}_{(1)} = \overline{x}_1 = 9'06 \\ \overline{x}_{(2)} = \overline{x}_4 = 7'98 \\ \overline{x}_{(3)} = \overline{x}_2 = 5'76 \\ \overline{x}_{(4)} = \overline{x}_3 = 5'48 \\ \overline{x}_{(5)} = \overline{x}_5 = 4'94 \end{array}$$

En cuanto al valor de MC_E*, se obtiene*

$$MC_E = 0'3461.$$

Se trata ahora de ir comparando cada dos medias con el valor

$$R_p = r_p\sqrt{\frac{MC_E}{n}},$$

donde r_p *viene dado en la tabla A.44.*

$$\begin{aligned} R_p &= r_p\sqrt{\frac{MC_E}{n}} \\ &= r_p\sqrt{\frac{0'3461}{5}} = r_p \cdot 0'2631. \end{aligned}$$

Buscando en las tablas para $n = 20$, se obtiene que

$$r_5 = 3'25;\ r_4 = 3'18;\ r_3 = 3'10;\ r_2 = 2'95,$$

por lo que

$$\begin{array}{ll} R_5 = 0'8551; & R_4 = 0'8367; \\ R_3 = 0'8156; & R_2 = 0'7761. \end{array}$$

Para comparar las medias correspondientes a Pastillas y Yoga, grupos 1 y 2, se procede en primer lugar a identificar los rangos correspondientes a dichas medias, resultando que $\overline{x}_1 = \overline{x}_{(1)}$ y $\overline{x}_2 = \overline{x}_{(3)}$. A continuación se calcula la diferencia de las medias

$$\overline{x}_{(1)} - \overline{x}_{(3)} = 9'06 - 5'76 = 3'3$$

el número de medias a considerar es $p = (3 - 1) + 1 = 3$, por lo que dicha diferencia debe compararse con $R_3 = 0'8156$, resultando significativa.

*El resultado de las restantes comparaciones se recoge en la tabla 6.2, donde se han marcado con * las diferencias significativas.*

Resumiendo, las pastillas y las hierbas presentan diferencias con los restantes grupos, pudiendo considerarse iguales las parejas Yoga-Autosugestión y Autosugestión-Gimnasia.

3.2. Test de Newman–Keuls

Este test es similar al de Duncan, salvo que el valor crítico se obtiene de modo diferente, a partir de las tablas de recorrido studentizado. Al igual que en el método anterior, se requiere que el número de observaciones por nivel sea el mismo, n.

Las k medias muestrales se disponen en orden creciente. De nuevo se van calculando las diferencias entre cada dos de ellas y se estudia si

	Rangos	Diferencia	p	R_p	
Pastillas-Yoga	(1)-(3)	3’3	3	0’8156	*
Pastillas-Autosugestión	(1)-(4)	3’58	4	0’8367	*
Pastillas-Hierbas	(1)-(2)	1’08	2	0’7761	*
Pastillas-Gimnasia	(1)-(5)	4’12	5	0’8551	*
Yoga-Autosugestión	(3)-(4)	0’28	2	0’7761	
Yoga-Hierbas	(3)-(2)	2’22	2	0’7761	*
Yoga-Gimnasia	(3)-(5)	0’82	3	0’8156	*
Autosugestión-Hierbas	(4)-(2)	2’5	3	0’8156	*
Autosugestión-Gimnasia	(4)-(5)	0’54	2	0’7761	
Hierbas-Gimnasia	(2)-(5)	3’04	4	0’8367	*

Tabla 6.2: Resultados: Test de Duncan

son mayores que los valores

$$W_p = q_{\alpha,p,N-k}\sqrt{\frac{MC_E}{n}},$$

siendo p el número de medias que hay entre dos de ellas (incluidas éstas) y $q_{\alpha,p,N-k}$ el valor crítico dado en la tabla A.42.

Si alguna de esas diferencias es mayor que su correspondiente W_p, se rechaza que las medias sean iguales; en caso contrario, no se puede rechazar que sean iguales.

Ejemplo 6.8 *Para los datos del ejemplo anterior, se calcula*

$$W_p = q_{\alpha,p,N-k}\sqrt{\frac{MC_E}{n}},$$

donde

$$q_{\alpha,p,N-k} = q_{0'05,p,20}, \qquad p = 2,\ldots,5,$$

obteniéndose

$$q_5 = 4'23;\ q_4 = 3'96;\ q_3 = 3'58;\ q_2 = 2'95.$$

Por tanto,

$$W_5 = 1'1155; \qquad W_4 = 1'0419;$$
$$W_3 = 0'9419; \qquad W_2 = 0'7761.$$

A continuación se van comparando medias.

Medias 1 y 5: $\overline{x}_1 - \overline{x}_5 = 4'12 > W_5 = 1'1155$, *por tanto, se rechaza la hipótesis de igualdad de las medias 1 y 5.*

Comparando todas las medias, se llega a los mismos resultados que con Duncan, salvo que para el caso de las medias 2 y 5: $\overline{x}_2 - \overline{x}_5 = 0'82 < W_3 = 0'9412$, *luego no se puede rechazar la hipótesis de igualdad de ambas medias. Nótese que con el método anterior se rechazaba pero por una diferencia muy pequeña.*

Por tanto, los resultados son básicamente los mismos que en el ejemplo anterior.

3.3. Test de Bonferroni

El test de Bonferroni para comparaciones múltiples está basado en la distribución t de Student. Al contrario que sucedía en los test anteriores, aquí no es necesario que el número de observaciones por nivel sea el mismo, pero requiere la especificación previa del número total de estimaciones a realizar, m. A su vez, se trata de un procedimiento no secuencial, recurriendo a un único valor crítico para todas las diferencias.

La expresión que proporciona los intervalos de confianza para las m combinaciones lineales, L_i, siendo para todos ellos el nivel de confianza $1 - \alpha$, viene dada por

$$\hat{L}_i \pm t_{1-\frac{\alpha}{2m},N-k}\sqrt{\mathrm{V}[\hat{L}_i]}.$$

En el caso, $L_i = \mu_i - \mu_j$, las estimaciones por intervalos de confianza de Bonferroni para las diferencias en los valores medios de los tratamientos, a un nivel $1-\alpha$, vienen dadas por

$$\overline{x}_{i.} - \overline{x}_{j.} \pm t_{1-\frac{\alpha}{2m},N-k}\sqrt{MC_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

y por tanto, se rechazará la hipótesis de igualdad de las medias i y j cuando

$$|\overline{x}_{i.} - \overline{x}_{j.}| > t_{1-\frac{\alpha}{2m},N-k}\sqrt{MC_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}.$$

3.4. Test HSD de Tukey

El test HSD de Tukey, al igual que sucedía con el anterior, utiliza un único valor crítico para todas las diferencias posibles, en vez de comparar cada media con un valor distinto que dependa del número de medias contenidas en la comparación. En este sentido, este test es no secuencial. En concreto, el valor crítico utilizado es el que aparece en las tablas de recorrido studentizado para $p=k$, $q_{\alpha,k,N-k}$.

Para determinar si los valores medios de dos tratamientos difieren o no, sean por ejemplo μ_i y μ_j, con $i \neq j$, se determina el error estándar de la diferencia de medias entre las muestras. Este es

$$E_{i,j} = \sqrt{MC_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}.$$

A partir de él se construye el valor

$$HSD_{i,j} = \frac{q_{\alpha,k,N-k}}{\sqrt{2}}E_{i,j},$$

donde el valor $q_{\alpha,k,N-k}$ viene dado en la tabla A.42. Se considerará que las medias μ_i y μ_j son distintas cuando

$$|d_{i,j}| = |\overline{x}_{i.} - \overline{x}_{j.}| > HSD_{i,j}.$$

Este estadístico permite construir intervalos de confianza para todas las diferencias de medias, con un nivel de confianza de al menos $1-\alpha$, que viene dado por:

$$\overline{x}_{i.} - \overline{x}_{j.} \pm HSD_{i,j}.$$

Cuando todos los tamaños muestrales son iguales, el nivel de confianza por el método de Tukey es exactamente $1-\alpha$. Si los tamaños muestrales son diferentes, el nivel de confianza para todas las comparaciones por parejas será mayor, es decir, el método es conservador, proporcionando estimaciones menos precisas.

Ejemplo 6.9 *Si se le aplica el test HSD de Tukey a los datos del ejemplo 6.7, donde*

$$n_i = 5, \qquad \forall i \in \{1, \ldots, 5\},$$

se obtiene lo siguiente:

El valor de $q_{\alpha,k,N-k} = q_{0'05,5,20} = 4'23$ se encuentra en la tabla A.42. A partir de éste se calcula

$$HSD_{i,j} = \frac{q_{0'05,5,20}}{\sqrt{2}} E_{i,j}.$$

Como los tamaños muestrales son todos iguales, el valor de $E_{i,j}$ es constante, siendo éste,

$$\begin{aligned} E = E_{i,j} &= \sqrt{MC_E \frac{2}{5}} \\ &= 0'2631\sqrt{2}, \end{aligned}$$

luego

$$\begin{aligned} HSD = HSD_{i,j} &= q_{0'05,5,20} \cdot 0'2631 \\ &= 4'23 \cdot 0'2631 = 1'1155, \end{aligned}$$

que es constante para todas las medias.

Si se realizan los contrastes de igualdad de cada dos medias al igual que se hizo con Duncan y Newman–Keuls, se sacan básicamente las mismas conclusiones, salvo en un caso: el método de las pastillas y el de las hierbas se pueden considerar iguales según el test de Tukey.

3.5. Test de Scheffé

Hay ocasiones en las que no se sabe de antemano cuántos contrastes se desean realizar, o situaciones en las que interesa realizar más de

$k-1$ posibles comparaciones. En estos casos se recurre al método de Scheffé, que permite realizar cualquier contraste, incluso aquellos que plantean combinaciones lineales de medias de los tratamientos. Para la realización de este test, no es necesario que el número de observaciones por nivel sea el mismo.

Partiendo del modelo ANOVA:

$$X_i = \mu + \delta_i + \varepsilon_i \qquad \begin{cases} i = 1, \ldots, k \\ \varepsilon_i \sim N(0, \sigma) \end{cases}$$

se define *Contraste* como

$$L = \sum_{i=1}^{k} c_i \delta_i, \quad \text{con } \sum_{i=1}^{k} c_i = 0,$$

que no es sino una combinación lineal de los k efectos. Como puede observarse, este contraste equivale a

$$L = \sum_{i=1}^{k} c_i \mu_i, \quad \text{con } \sum_{i=1}^{k} c_i = 0,$$

ya que $\mu_i = \mu + \delta_i$ y $\sum_{i=1}^{k} c_i = 0$.

Puesto que L es desconocido, se puede obtener una estimación de la expresión anterior sustituyendo μ_i por su estimador, $\overline{x}_{i\cdot}$. Así pues, se tiene que

$$\hat{L} = \sum_{i=1}^{k} c_i \overline{x}_{i\cdot}.$$

En cuanto a la varianza de $\hat{L}$, se comprueba que vale

$$\sigma_{\hat{L}}^2 = \mathrm{V}[\hat{L}] = \sum_{i=1}^{k} c_i^2 \, \mathrm{V}[\overline{x}_{i\cdot}] = \sigma^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i},$$

y su estimación

$$\hat{\sigma}_{\hat{L}}^2 = MC_E \sum_{i=1}^{k} \frac{c_i^2}{n_i},$$

siendo MC_E la estimación de σ^2.

Así pues, el intervalo de confianza para L, a un nivel $1-\alpha$, viene dado por

$$\hat{L} \pm \hat{\sigma}_{\hat{L}} \sqrt{(k-1)F_{1-\alpha,k-1,N-k}}.$$

Esto se traduce en lo siguiente. Si se pretende probar la hipótesis de que el contraste L difiere significativamente de cero, la regla de decisión que se adopta es si $|\hat{L}| > \hat{\sigma}_{\hat{L}} \sqrt{(k-1)F_{1-\alpha,k-1,N-k}}$, se rechaza la hipótesis de que el contraste L es igual a cero.

Puesto que lo que ha motivado el planteamiento de los contrastes múltiples es la necesidad de averiguar si un nivel del factor presenta un efecto mayor que el de otro, se trata ahora de particularizar este método al estudio de contraste de parejas de medias.

Si se quiere contrastar

$$\begin{cases} H_0 : \mu_i = \mu_j \\ H_1 : \mu_i \neq \mu_j \end{cases}$$

la diferencia $L = \mu_i - \mu_j$ es un contraste de los definidos anteriormente, ya que la suma de sus coeficientes da cero. Averiguar si las medias i y j difieren es estudiar si la diferencia L es significativamente distinta de cero. Por tanto, bastará comprobar si

$$|\hat{L}| > \hat{\sigma}_{\hat{L}} \sqrt{(k-1)F_{1-\alpha,k-1,N-k}},$$

es decir, si

$$|\overline{x}_{i.} - \overline{x}_{j.}| > \hat{\sigma}_{\hat{L}} \sqrt{(k-1)F_{1-\alpha,k-1,N-k}}.$$

De ser así, se rechazaría la hipótesis de igualdad de medias.

Ejemplo 6.10 *Se quiere estudiar el grado de cultura que tienen los jóvenes entre 14 y 17 años según el nivel socio–económico, que podría clasificarse en 5 grupos: A, B, C, D y E. Aleatoriamente se toman muestras de 6 jóvenes pertenecientes a cada nivel y se les somete a un test, obteniendo los siguientes resultados:*

A	B	C	D	E
32	37	30	25	39
35	39	35	30	42
27	30	28	22	28
31	33	34	27	33
38	42	34	31	47
29	28	30	21	30

Suponiendo que entre los cinco niveles existen diferencias significativas, se trata de averiguar entre qué grupos se dan las diferencias, para lo que se recurre al test de Scheffé.

En primer lugar, se quiere realizar el contraste

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Se trata de averiguar si

$$|\overline{x}_{1.} - \overline{x}_{2.}| \quad > \quad \hat{\sigma}_{\hat{L}} \sqrt{(k-1)F_{1-\alpha,k-1,N-k}},$$

para el nivel de significación $0'05$. *Esto es equivalente a*

$$\frac{(\overline{x}_{1.}-\overline{x}_{2.})^2}{\hat{\sigma}^2_{\hat{L}}} \frac{1}{k-1} \quad > \quad F_{1-\alpha,k-1,N-k}.$$

La tabla siguiente recoge los datos necesarios para aplicar el test:

	x	x^2	n	$\overline{x}_{i.}$
A	192	6224	6	32
B	209	7427	6	$34'83$
C	191	6121	6	$31'83$
D	156	4140	6	26
E	219	8267	6	$36'5$
Total	967	32179	30	$32'23$

$$MC_E = \frac{SC_E}{N-k} = \frac{625'17}{25} = 25'01.$$

Se calcula

H_0	F_e
$\mu_1 = \mu_2$	$0'24$
$\mu_1 = \mu_3$	$0'00$
$\mu_1 = \mu_4$	$1'08$
$\mu_1 = \mu_5$	$0'61$
$\mu_2 = \mu_3$	$0'27$
$\mu_2 = \mu_4$	$0'34$
$\mu_2 = \mu_5$	$0'08$
$\mu_3 = \mu_4$	$1'02$
$\mu_3 = \mu_5$	$0'65$
$\mu_4 = \mu_5$	$3'31^*$

Tabla 6.3: Resultados del test de Scheffé

$$\frac{(\overline{x}_{1.} - \overline{x}_{2.})^2}{\hat{\sigma}^2_{\hat{L}}}\frac{1}{k-1} \quad = \quad \frac{(-2'83)^2}{25'01(\frac{1}{6}+\frac{1}{6})}\frac{1}{4} = 0'24.$$

Como $F_{1-\alpha,k-1,N-k} = F_{0'95,4,25} = 2'76 < 0'24$, no se puede rechazar la hipótesis de igualdad de las medias 1 y 2, es decir, que no existen diferencias significativas entre el grupo A y B.

*Habría que seguir realizando los contrastes de pares de medias hasta comparar todas. Los resultados obtenidos se recogen en la tabla 6.3, donde se han marcado con * las diferencias significativas.*

3.6. Diferencias entre los test de comparaciones múltiples

Dependiendo de que las comparaciones sean entre parejas de tratamientos o más generales, será más aconsejable el test de Tukey o el de Scheffé. En el caso de comparaciones de parejas de tratamientos, puesto que el de Tukey proporciona intervalos de confianza de menor longitud, se preferirá al de Scheffé. Sin embargo, puesto que normalmente los contrastes que se realizan son más generales, el test de Scheffé tiende a proporcionar, en esos casos, intervalos de confianza de menor longitud.

Si el número de contrastes a realizar es, a lo sumo, el número de tratamientos, es preferible el test de Bonferroni al de Scheffé. De hecho, para que el test de Scheffé sea mejor, el número de contrastes debe ser bastante superior al de tratamientos o niveles del factor.

Por otra parte, en lo que respecta a la elección entre Duncan, Newman–Keuls o Tukey, todo depende del riesgo que se esté dispuesto a correr al aceptar más o menos diferencias entre las medias. Se elegirá entre uno u otro según se esté interesado en un menor o mayor número de diferencias significativas y el grado de seguridad con que pueda tomarse.

4. Validación del modelo

En el modelo ANOVA de un factor,

$$X_i = \mu + \delta_i + \varepsilon_i, \quad \forall i = 1, \ldots, k$$

se parte del supuesto que $\varepsilon_i \sim N(0, \sigma)$ y de que las muestras son independientes. Cuando estas suposiciones son válidas, el modelo proporciona una prueba exacta para la hipótesis de igualdad de medias de los niveles. Pero en la práctica, éstas no se suelen cumplir con exactitud, por lo que antes de aplicar el análisis de la varianza deberá comprobarse que tales suposiciones son ciertas.

Para averiguar si se cumplen las mencionadas condiciones paramétricas de aleatoriedad, normalidad y homocedasticidad de los residuos o errores, se hace uso de algunos de los contrastes vistos anteriormente.

Así, para estudiar la aleatoriedad de los ε_i, se puede recurrir al test de rachas. En cuanto al estudio de la normalidad de los residuos, lo mejor es realizar uno de los siguientes contrastes: el test de la χ^2, el test de Kolmogorov–Smirnov–Lilliefors o el test W de Shapiro–Wilks.

4.1. Homocedasticidad

Para estudiar si los residuos provienen de poblaciones Normales de igual varianza, es decir para comprobar la homocedasticidad, existen varios test que contrastan dicha igualdad de varianzas en k poblaciones.

Los contrastes de Bartlett y Hartley, que suponen que cada una de las k poblaciones es Normal y que se han obtenido de cada población muestras aleatorias simples, o el de Cochran, son algunos de los que se proponen.

El de Bartlett es un contraste general que abarca tanto el caso de tamaños muestrales iguales como desiguales; sin embargo, el de Hartley sólo es aplicable si los tamaños muestrales son iguales, y se ha diseñado para ser sensible a diferencias sustanciales entre la varianza poblacional más grande y la más pequeña. El contraste de Cochran se preferirá cuando la mayor varianza muestral sea mucho mayor que el resto, o cuando el número de niveles por factor sea superior a 12.

En los tres, el contraste que se pretende realizar es el siguiente:

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_k \\ H_1 : \sigma_i \neq \sigma_j, \text{para algún } i \neq j \end{cases}$$

4.1.1. Test de Bartlett

Para resolver el contraste de igualdad de varianzas propuesto, se recurre al siguiente estadístico:

$$B = \frac{1}{C}\left[(N-k)\log S_c^2 - \left(\sum_{i=1}^{k}(n_i-1)\log S_{c_i}^2\right)\right],$$

donde $S_{c_1}^2, \ldots, S_{c_k}^2$ representan las cuasivarianzas muestrales, S_c^2 denota la cuasivarianza ponderada,

$$S_c^2 = \frac{\sum_{i=1}^{k}(n_i-1)S_{c_i}^2}{N-k},$$

y donde C viene dada por

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right].$$

Para un nivel de significación α, el test de Bartlett rechaza H_0, si $B > \chi^2_{k-1,1-\alpha}$.

Observación 6.1 *La aproximación a una Chi–cuadrado se considera apropiada cuando* $n_i \geq 5$, $i = 1, \ldots, k$.

El test de Bartlett es sensible a desviaciones de la normalidad, en cuyo caso no es recomendable.

Ejemplo 6.11 *Se está realizando un estudio para averiguar cuánto difieren los niveles de una determinada sustancia química en tres grupos de personas que han seguido dietas diferentes. Para ello se tomaron las siguientes muestras, que se saben están normalmente distribuidas:*

Dieta 1	*Dieta 2*	*Dieta 3*
256	266	269
628	256	256
253	258	620
256	320	452
	450	286
		256

Con el propósito de averiguar en estudios posteriores si el nivel de la sustancia depende de la dieta, se va a llevar a cabo un análisis del cumplimiento de la condición de homocedasticidad, a través del test de Bartlett. El contraste que se plantea es

$$\left\{ \begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \\ H_1 : \sigma_i^2 \neq \sigma_j^2, \text{ para algún } i \neq j \end{array} \right.$$

El estadístico que se utiliza es

$$B = \frac{1}{C} \left[(N-k) \log S_c^2 - \left(\sum_{i=1}^{k} (n_i - 1) \log S_{c_i}^2 \right) \right]$$

siendo

$$S_c^2 = \frac{\sum_{i=1}^{k} (n_i - 1) S_{c_i}^2}{N-k}$$

y

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right].$$

Es sabido que $B \sim \chi^2_{k-1}$. *Para* $k = 3$ *y puesto que* $N = 15$, $n_1 = 4$, $n_2 = 5$ *y* $n_3 = 6$, *el valor de* C *será,*

$$\begin{aligned} C &= 1 + \tfrac{1}{3\cdot 2}\left[\tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{12}\right] \\ &= 1'1167. \end{aligned}$$

Calculando las cuasivarianzas muestrales,

$$\begin{aligned} S^2_{c_i} &= \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i)^2 \\ &= \frac{1}{n_i-1}\left[\sum_{j=1}^{n_i}x_{ij}^2 - \frac{1}{n_i}\left(\sum_{j=1}^{n_i}x_{ij}\right)^2\right] \end{aligned}$$

$$S^2_{c_1} = 34784'25; S^2_{c_2} = 6814; S^2_{c_3} = 22275'9$$

y el valor de S^2_c *resulta*

$$\begin{aligned} S^2_c &= \frac{1}{12}\left[3\cdot 34784'25 + 4\cdot 6814 + 5\cdot 22275'9\right] \\ &= \frac{242988'25}{12} = 20249'0208. \end{aligned}$$

Luego, sustituyendo en la expresión de B, *se llega a que* $B = 2'0206$. *Para un nivel de significación de* $0'05$, *el valor de* $\chi^2_{2,0'95} = 5'99$, *que resulta ser mayor que el de* B, *por lo que no se puede rechazar la hipótesis de igualdad de varianzas, asumiendo por tanto, que se verifica la hipótesis de homocedasticidad.*

4.1.2. Test de Cochran

El test de Cochran para muestras del mismo tamaño, n, propone como estadístico la razón entre la mayor varianza muestral y la suma de todas las varianzas muestrales, es decir,

$$C = \frac{\text{máx}_{1\leq i\leq k} S_i^2}{\sum_{i=1}^{k} S_i^2}.$$

En la tabla A.39, aparecen los valores críticos para este contraste. Si resulta que C es mayor que el valor que allí apareciese para k y n, se rechaza la hipótesis de igualdad de varianzas.

Ejemplo 6.12 *Supóngase que se está estudiando un factor que tiene 5 niveles, de los cuales se extraen cinco muestras aleatorias de tamaño 10. Se conocen las varianzas de cada una de las muestras:*

$$S_1^2 = 26, S_2^2 = 51, S_3^2 = 40, S_4^2 = 24 \text{ y } S_5^2 = 28.$$

Se trata de realizar el contraste de igualdad de varianzas, al nivel de significación de $0'05$, para las muestras anteriores, a través del test de Cochran.

El estadístico de contraste viene dado por:

$$\begin{aligned} C &= \frac{\displaystyle\max_{1\leq i\leq k} S_i^2}{\displaystyle\sum_{i=1}^{k} S_i^2} = \frac{51}{169} \\ &= 0'302. \end{aligned}$$

Buscando en las tablas de Cochran para $\alpha = 0'05$, $k = 5$ y $n = 10$, se obtiene el valor crítico de $0'4241$, que resulta ser mayor que el valor de C, por lo que no se puede rechazar la hipótesis de igualdad de varianzas.

4.1.3. Test de Hartley

El test de Hartley se aplica cuando las k varianzas muestrales tienen los mismos grados de libertad, es decir, cuando el número de observaciones por nivel sea el mismo, n. Está basado en la comparación de la mayor varianza muestral con la menor, de modo que el estadístico que se emplea es

$$H = \frac{\text{máx}_{1\leq i\leq k} S_i^2}{\text{mín}_{1\leq i\leq k} S_i^2}.$$

Cuando el estadístico H sea mayor que el valor tabulado $H_{k,\nu}$, siendo $\nu = n-1$, se rechaza la hipótesis de igualdad de las varianzas. Los valores de $H_{k,\nu}$ vienen dados en la tabla A.45.

4.2. Transformaciones de las observaciones

Hay ocasiones en las que una transformación en las observaciones permite corregir alguna de las desviaciones debidas al no cumplimiento de las hipótesis del modelo. Es el caso de los datos que, sin proceder de poblaciones Normales, pueden aproximarse a ella.

Entre las transformaciones más importantes destacan la transformación raíz cuadrada y la transformación arco seno.

4.2.1. Transformación raíz cuadrada

Esta transformación resulta muy útil cuando los datos proceden de una Poisson o, de modo general, cuando los datos observados son número de elementos. En estos casos se trabaja con la raíz cuadrada de los datos o con alguna función de la misma, como es

$$\sqrt{x + \frac{3}{8}}.$$

4.2.2. Transformación arco seno

Cuando los datos vienen dados en tantos por ciento, p, es decir, proceden de una distribución Binomial, se aconseja realizar la transformación

$$\operatorname{arc\,sen} \sqrt{p}.$$

Ejemplo 6.13 *Se quiere estudiar si los tres niveles de un determinado factor A son iguales. Para ello, se toman tres observaciones de cada nivel, viniendo dados los datos en porcentajes, tal y como indica la tabla siguiente:*

Nivel I	*Nivel II*	*Nivel III*
$8'1$	$8'6$	12
$9'2$	$8'9$	$13'2$
$9'5$	$7'4$	$13'1$

Previo al análisis de la varianza y puesto que los datos siguen una Binomial, se va a realizar una transformación de las observaciones mediante $\operatorname{arc\,sen}\sqrt{p}$. *Los resultados aparecen en la siguiente tabla:*

Nivel I	*Nivel II*	*Nivel III*
$16'54$	$17'05$	$20'27$
$17'66$	$17'36$	$21'30$
$17'95$	$15'79$	$21'22$

A continuación se realiza el correspondiente análisis de la varianza:

Fuente	*SC*	*G.L.*	*MC*	F_0
Entre	$30'61$	2	$15'31$	
Dentro	$3'15$	6	$0'52$	$29'172$

Como $F_{1-\alpha,k-1,N-k} = F(0'1,2,6) = 3'46$ *y*

$$F_0 = 29'172 \quad > \quad 3'46 = F_{1-\alpha,k-1,N-k}$$

se rechaza la hipótesis de que los tres niveles del factor sean iguales.

5. Tests no paramétricos

Cuando alguna de las condiciones paramétricas del ANOVA no se cumpla, habrá que recurrir a otros métodos para contrastar la hipótesis de igualdad en los niveles medios del factor que se está estudiando. Los tests no paramétricos son una alternativa al ANOVA cuando alguna de las condiciones de normalidad, homocedasticidad o independencia falla. Entre ellos, se estudian los test de Kruskal–Wallis, de Q de Cochran y el de Welch.

Una vez contrastada la hipótesis de igualdad de los efectos, si ésta es rechazada, habrá que realizar un estudio para averiguar donde se encuentran las diferencias. Para ello, se recurre de nuevo a las comparaciones múltiples, en el que se estudia el test de comparaciones múltiples de Kruskal–Wallis.

5.1. Test de Kruskal–Wallis

El test de Kruskal–Wallis permite probar la hipótesis de igualdad de k tratamientos cuando no se da la normalidad o la homocedasticidad, siendo necesario sólo que las observaciones sean independientes.

Dicho test se basa en reemplazar las observaciones por sus respectivos rangos, para lo cual se toman k muestras aleatorias de tamaño n_i de cada uno de los niveles del factor. A continuación se organizan las observaciones x_{ij} en orden ascendente, reemplazándolas por su rango, R_{ij}.

Si el número total de datos es $N = \sum_{i=1}^{k} n_i$, el rango 1 le corresponderá al valor más pequeño y al más grande, N. Cuando haya observaciones repetidas, se les asignará el rango promedio a cada uno de ellas. La suma de los rangos de las observaciones del i-ésimo nivel se denota por $R_{i.}$.

Cuando no existen observaciones repetidas, el estadístico de contraste tiene la siguiente expresión

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_{i.}^2}{n_i} - 3(N+1).$$

Cuando, por el contrario, existan varias observaciones repetidas, se utiliza una corrección del estadístico anterior

$$H_c = \frac{H}{C},$$

siendo

$$C = 1 - \frac{\sum_{i=1}^{k}(r_i^3 - r_i)}{N^3 - N},$$

donde r_i denota el número de observaciones repetidas en cada grupo de repeticiones.

Cuando $n_i \geq 5$ o el número de niveles es mayor que 3, el estadístico H se distribuye como una Chi–cuadrado de $k-1$ grados de libertad, si la Hipótesis Nula es verdadera. Por tanto, se rechaza la Hipótesis Nula cuando $H > \chi^2_{k-1,1-\alpha}$.

Para el caso de 3 niveles y $n_i \leq 5$, se recurre a los valores que aparecen en la tabla A.40, $H(\alpha, 3, (n_1, n_2, n_3))$, de manera que se rechazará la Hipótesis Nula cuando $H > H(\alpha, 3, (n_1, n_2, n_3))$.

Ejemplo 6.14 *Una compañía quiere comparar 5 tipos de llantas, A, B, C, D y E, por lo que decide probarlas en cinco coches similares, asignándoles a cada uno de ellos y al azar, un tipo de llanta. Sus vidas medias en rodaje, medidas en miles de millas, vienen dadas en la siguiente tabla:*

Llantas	*Vidas medias*				
A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Se quiere contrastar la hipótesis de que no hay diferencias entre ellas a un nivel de significación del 5 %.

Suponiendo que no se verifican las condiciones paramétricas, el problema se va a abordar desde el punto de vista no paramétrico, recurriendo al test de Kruskal–Wallis.

Para ello se ordenan los datos en orden creciente y se les asigna su rango, obteniéndose

Llantas	Rangos					R_i
A	$17'5$	21	24	1	$6'5$	70
B	21	$6'5$	12	$6'5$	$2'5$	$48'5$
C	10	25	14	23	21	93
D	$2'5$	11	9	14	4	$40'5$
E	14	16	19	$17'5$	$6'5$	73

Se tiene $N = 25$ observaciones, de las cuales algunas están repetidas. La siguiente tabla refleja aquellos datos que se repiten y cuantas veces:

Observación	48	53	64	68	72	*Total*
r_i	2	4	3	2	3	
$r_i^3 - r_i$	6	60	24	6	24	120

Puesto que existen observaciones repetidas, el estadístico de contraste es $H_c = \frac{H}{C}$, siendo

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_{i.}^2}{n_i} - 3(N+1)$$

$$C = 1 - \frac{\sum_{i=1}^{k} (r_i^3 - r_i)}{N^3 - N}.$$

A partir de las tablas anteriores, se obtiene

$$C = 0'9923 \quad y \quad H = 6'44$$

Por tanto, el valor del estadístico

$$H_c = \frac{6'44}{0'9923} = 6'49.$$

Para $\alpha = 0'05$ es $\chi^2_{k-1,1-\alpha} = \chi^2_{4,0'95} = 9'49$, resultando ser mayor que el valor H_c, por lo que no se puede rechazar la hipótesis de igualdad en las llantas.

5.2. Test Q de Cochran

El test de Cochran se utiliza cuando las observaciones no cumplen ninguna de las condiciones paramétricas, en especial la independencia. Es el test que se suele aplicar cuando, por ejemplo, la misma persona o grupo de personas es sometida a los k niveles del factor o cuando los valores observados son dicotómicos. Es importante señalar que en este test, el número de elementos sometidos a cada nivel del factor debe ser el mismo.

La base radica en la construcción de una tabla de ceros y unos, con tantas filas como individuos sometidos a cada nivel (supóngase N) y tantas columnas como niveles tenga el factor (k). Cuando la respuesta del individuo al nivel sea favorable, se le asignará un uno, en caso contrario, un cero. El total de unos en la columna j se denotará por c_j y el de la fila i por f_i.

El estadístico de contraste es el siguiente:

$$Q = \frac{(k-1)\left[k\displaystyle\sum_{j=1}^{k} c_j^2 - \left(\sum_{j=1}^{k} c_j\right)^2\right]}{k\displaystyle\sum_{i=1}^{N} f_i - \sum_{i=1}^{N} f_i^2},$$

que se sabe tiene una distribución Chi–cuadrado con $k-1$ grados de libertad.

Ejemplo 6.15 *Una empresa está a punto de sacar al mercado un nuevo producto dirigido a los jóvenes entre 14 y 17 años, pero aún no ha decidido la apariencia y textura que le van a dar a dicho producto. Cuatro agencias de publicidad le han presentado sus propuestas, pero la agencia ha decidido que sean los propios consumidores los que elijan la forma. Aleatoriamente seleccionaron a seis jóvenes y les presentaron las cuatro modelos, teniendo que contestar si les gustaba o no. La siguiente tabla recoge* *los resultados, donde los unos significa que gustó y los ceros, que no.*

Joven	*Modelo 1*	*Modelo 2*	*Modelo 3*	*Modelo 4*
1	0	0	1	0
2	0	0	1	1
3	0	1	1	1
4	1	1	0	0
5	1	0	0	1
6	0	1	1	0

Para contrastar si los cuatro modelos son igualmente aceptados, se recurre al test Q de Cochran.

El estadístico de contraste es,

$$Q = \frac{(k-1)\left[k\sum_{j=1}^{k} c_j^2 - \left(\sum_{j=1}^{k} c_j\right)^2\right]}{k\sum_{i=1}^{N} f_i - \sum_{i=1}^{N} f_i^2},$$

por lo que hay que calcular las f_i y c_i. Observando la tabla, se obtiene:

$$c_1 = 2; c_2 = 3; c_3 = 4; c_4 = 3;$$

$$f_1 = 1; f_2 = 2; f_3 = 3; f_4 = 2; f_5 = 2; f_6 = 2;$$

$$\textstyle\sum_{j=1}^{4} c_j^2 = 38; \qquad \left(\sum_{j=1}^{4} c_j\right)^2 = 144$$

$$\textstyle\sum_{i=1}^{6} f_i = 12; \qquad \sum_{i=1}^{6} f_i^2 = 26.$$

Por lo tanto,

$$Q = \frac{3[4 \cdot 38 - 144]}{4 \cdot 12 - 26} = 1'091$$

y puesto que, para un nivel de significación del 10 %, es $\chi_3^2 = 6'251$, mayor que Q, no se puede rechazar la hipótesis de igualdad en los cuatro modelos propuestos.

5.3. Test de Welch

Para contrastar si los efectos de los distintos niveles de un factor son iguales, cuando falla la condición de homocedasticidad, se utiliza el test de Welch.

Se extraen k muestras aleatorias de cada nivel, de tamaño n_i. Sea $w_i = \frac{n_i}{S_i^2}$, siendo S_i^2 la varianza del nivel i-ésimo, y $w = \sum_{i=1}^k w_i$. Si $\overline{x}_{i.}$ representa la media del nivel i-ésimo, se define

$$\overline{x}_{..} = \frac{\displaystyle\sum_{i=1}^k w_i \overline{x}_{i.}}{\displaystyle\sum_{i=1}^k w_i}.$$

El estadístico que se utiliza es el siguiente:

$$V^2 = \frac{\dfrac{\displaystyle\sum_{i=1}^k w_i(\overline{x}_{i.} - \overline{x}_{..})^2}{k-1}}{1 + \dfrac{2(k-2)}{k^2-1}\displaystyle\sum_{i=1}^k \frac{1}{n_i - 1}\left[1 - \frac{w_i}{w}\right]^2},$$

que se sabe tiene una distribución F de Snedecor de ν_1 y ν_2 grados de libertad, siendo

$$\nu_1 = k - 1; \quad \nu_2 = \left[\frac{3}{k^2-1}\sum_{i=1}^k \frac{1}{n_i - 1}\left(1 - \frac{w_i}{w}\right)^2\right]^{-1}.$$

Si $V^2 > F_{\nu_1,\nu_2,\alpha}$, se rechaza la hipótesis de que las los niveles medios del factor sean iguales.

Ejemplo 6.16 *En una academia de francés, disponen de tres métodos para el aprendizaje de este idioma. Se* *quiere averiguar si los tres son igualmente efectivos, para lo que se toman muestras aleatorias de individuos pertenecientes a cada uno de los grupos sometidos a un método. Tras hacerles un examen, los resultados fueron los siguientes:*

Método 1	*Método* 2	*Método* 3
$117'4$	$136'3$	$212'6$
$196'6$	$140'4$	$217'6$
$197'5$	$175'5$	$220'6$
$103'8$	$104'4$	$220'0$
	$175'3$	

Si se realiza el test de Shapiro–Wilk, a un nivel de significación del $10\,\%$*, para saber si se cumple la condición de normalidad, se llega a que* $W_{exp} = 0'879$ *y* $W_{13,0'1} = 0'889$*, con lo que se rechaza la hipótesis de normalidad. Análogamente, si se realiza el test de Hartley para contrastar la hipótesis de igualdad de varianzas, se llega a rechazar dicha hipótesis. Ante esta situación, para averiguar si los tres métodos son igualmente efectivos se va a recurrir al test de Welch.*

La siguiente tabla recoge los datos que hacen falta para calcular el estadístico de contraste:

	Método 1	*Método 2*	*Método 3*
n_i	4	5	4
$\overline{x}_{i.}$	$152'83$	$146'38$	$217'70$
S_i^2	$1891'6219$	$716'8056$	$9'93$
$w_i = \dfrac{n_i}{S_i^2}$	$0'0021$	$0'007$	$0'4028$

Sustituyendo los valores de

$$\begin{aligned} w &= \sum_{i=1}^{3} w_i = 0'41191; \\ \overline{x}_{..} &= \frac{\sum_{i=1}^{3} w_i \overline{x}_{i.}}{w} = 216'15909 \end{aligned}$$

en la expresión del estadístico, se obtiene,

$$\begin{aligned} V^2 &= \frac{\dfrac{\displaystyle\sum_{i=1}^{k} w_i(\overline{x}_{i.} - \overline{x}_{..})^2}{k-1}}{1 + \dfrac{2(k-2)}{k^2-1} \displaystyle\sum_{i=1}^{k} \frac{1}{n_i - 1} \left[1 - \frac{w_i}{w}\right]^2} \\ &= 18'987. \end{aligned}$$

Puesto que $\nu_1 = 2$ *y*

$$\begin{aligned} \nu_2 &= \left[\frac{3}{k^2-1} \sum_{i=1}^{k} \frac{1}{n_i - 1} \left(1 - \frac{w_i}{w}\right)^2\right]^{-1} \\ &= 4'66, \end{aligned}$$

para un nivel de significación de $0'1$*, se tiene que* $F_{2,5,0'1} = 3'78$ *y como* $V^2 > F_{2,5,0'1}$*, se rechaza la hipótesis de que los tres métodos sean iguales.*

5.4. Test de comparaciones múltiples de Kruskal–Wallis

Si tras realizar el contraste se rechaza la hipótesis de igualdad en los niveles medios del factor, habrá que averiguar cuál o cuáles de ellos son distintos.

Suponiendo, en primer lugar, que todas las muestras son del mismo tamaño, n, el criterio adoptado para rechazar la hipótesis de que la media de los niveles i y j sea la misma, viene dada por:

$$|R_i - R_j| \geq r_{\alpha,k,n},$$

siendo R_i la suma de los rangos en la muestra del nivel i, R_j la suma de los rangos de la muestra del nivel j, k el número de niveles del factor y $r_{\alpha,k,n}$ dado en la tabla A.41.

Cuando $k = 3$ y las muestras son de distinto tamaño, el criterio para rechazar la hipótesis de igualdad de medias de los niveles i y j, es el siguiente:

$$|\overline{R}_i - \overline{R}_j| \geq \sqrt{H(\alpha, 3, (n_1, n_2, n_3))\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

donde $\overline{R}_i$ y $\overline{R}_j$ son los rangos medios de las muestras i y j, es decir,

$$\overline{R}_i = \frac{R_i}{n_i} \quad \overline{R}_j = \frac{R_j}{n_j}$$

y $H(\alpha, 3, (n_1, n_2, n_3))$ viene dado en la tabla A.40.

Cuando se trate de muestras grandes y de distinto tamaño, se utiliza la aproximación

$$|\overline{R}_i - \overline{R}_j| \geq Z_\alpha \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

donde Z_α es el valor crítico de la distribución $N(0, 1)$ para un nivel de significación

$$\frac{\alpha}{k(k-1)}.$$

6. Ejercicios

6.1. Ejercicio resuelto

6.1 El propietario de una tienda de electrodomésticos tiene pensado asignar unas primas en función de la capacidad de venta de sus 3 vendedores. Para conocer dicha capacidad, registra, en cinco instantes distintos de tiempo el volumen de ventas realizado por cada uno. Dichas cantidades, en miles, vienen dadas en la siguiente tabla:

Vendedor A	Vendedor B	Vendedor C
30	85	40
20	73	28
35	92	39
42	86	41
60	75	50

a) Comprobar, a un nivel de confianza del 90 %, si el volumen de ventas difiere según el vendedor.

b) Suponiendo normalidad, independencia y homocedasticidad en los datos, ¿se puede admitir con una confianza del 99 % que el volumen de ventas depende del empleado?

c) ¿Se puede afirmar, con una confianza del 95 %, que la capacidad de venta del empleado A es diferente a la del empleado C?

Solución:

a) Para averiguar si el volumen de ventas difiere de unos empleados a otros, se plantea el siguiente contraste de medias:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 \\ H_1 : \mu_i \neq \mu_j, \text{ para algún } i, j \in \{1, 2, 3\} \end{cases}$$

Suponiendo que no se verifican las condiciones paramétricas, el problema se va a abordar desde el punto de vista no paramétrico, recurriendo al test de Kruskal–Wallis.

Lo primero es ordenar los datos y asignarles su rango, obteniendo la siguiente tabla:

A	r_A	B	r_B	C	r_C
30	3	85	13	40	6
20	1	73	11	28	2
35	4	92	15	39	5
42	8	86	14	41	7
60	10	75	12	50	9
$R_{1.}=$	26	$R_{2.}=$	65	$R_{3.}=$	29

Puesto que hay $N = 15$ datos y ninguno de ellos se repite, el estadístico de contraste viene dado por

$$\begin{aligned} H &= \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_{i.}^2}{n_i} - 3(N+1) \\ &= \frac{12}{15 \cdot 16} \left[\frac{26^2}{5} + \frac{65^2}{5} + \frac{29^2}{5} \right] - 3 \cdot 16 = 9'42. \end{aligned}$$

Utilizando la tabla A.40, para $\alpha = 0'1$ y $n_1 = n_2 = n_3 = 5$, se obtiene el valor $H(0'1, 3, (5, 5, 5)) = 4'56$. Como $H(\alpha, 3, (n_1, n_2, n_3)) < H$, se rechaza la hipótesis de igualdad de medias, es decir, existen evidencias para decir que la capacidad de venta difiere de unos empleados a otros.

b) Suponiendo ahora normalidad, homocedasticidad e independencia, para averiguar si el volumen de ventas depende del empleado, o lo que es lo mismo, si el volumen de ventas es el mismo para cada empleado, se recurre al test F del ANOVA.

De nuevo se plantea el contraste:

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 = \mu_3 \\ H_1 : \mu_i \neq \mu_j, \text{ para algún } i, j \in \{1, 2, 3\} \end{array} \right.$$

Para resolverlo, se recurre al estadístico

$$F = \frac{MC_{Tr}}{MC_E} \sim F_{k-1,N-k},$$

donde $k = 3$ y $N = 15$. La siguiente tabla recoge los datos que hacen falta para el cálculo de las medias de cuadrados, MC_{Tr} y MC_E.

Nivel	x_{ij}	n_i	$x_{i.}$	$\sum_j x_{ij}^2$	$x_{i.}^2/n_i$
A	30, 20, 35, 42, 60	5	187	7889	$6993'8$
B	85, 73, 92, 86, 75	5	411	34039	$33784'2$
C	40, 28, 39, 41, 50	5	198	8086	$7840'8$
Total		15	796	50014	$48618'8$

Una vez organizados los datos, se calcula la suma de cuadrados, SC_{Tr} y SC_E

$$SC_{Tr} = \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N} = 6377'7333$$

$$SC_E = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} = 1395'2.$$

Y la tabla ANOVA vendría dada por:

Fuente	SC	$G.L.$	MC	F_0
Entre	$6377'7333$	2	$\frac{6377'7333}{2} = 3188'8667$	
Dentro	$1395'2$	12	$\frac{1395'2}{12} = 116'2667$	$\frac{3188'8667}{116'2667} = 27'4272$
Total	$7772'9333$	14		

Para un nivel de significación de $0'01$, se tiene que $F_{2,12,0'01} = 6'93$, que es menor que $F_0 = 27'4272$, por lo que se rechaza la hipótesis de igualdad de medias. Es decir, que el volumen de ventas varía según el empleado.

c) Para determinar si la capacidad de venta del empleado A es diferentes de la del empleado C, se va a recurrir al método de Scheffé.

El contraste que se quiere realizar es el siguiente:

$$\begin{cases} H_0 : \mu_1 = \mu_3 \\ H_1 : \mu_1 \neq \mu_3 \end{cases}$$

El estadístico que se toma es

$$F = \frac{(\overline{X}_1 - \overline{X}_3)^2}{(k-1)\frac{SC_E}{n-k}\left(\frac{1}{n_1} + \frac{1}{n_3}\right)} \sim F_{k-1,n-k}.$$

Así,

$$F = \frac{(\frac{187}{5} - \frac{198}{5})^2}{(3-1)\frac{1395'2}{15-3}\left(\frac{1}{5} + \frac{1}{5}\right)} = 0'0520$$

y como, para el nivel de significación del 1 % es $F_{2,12} = 6'93 > 0'0520$, no se puede rechazar la hipótesis de igualdad de las medias 1 y 3, es decir, no existen evidencias para decir que la capacidad de ventas del empleado A y C sean diferentes, por lo que tendremos que asumir que son iguales.

6.2. Ejercicios propuestos

6.1. Se está realizando un estudio sobre la tolerancia y el respeto de los jóvenes en las aulas de tres tipos de colegio: público, concertado y privado. Tras tomar diferentes muestras aleatorias de cada uno de los colegios, los resultados fueron:

Público	Concertado	Privado
9	11	16
10	13	18
6	8	14
7	10	15
11	12	18
5	9	17
11	13	
12		

Estudie si existen diferencias en los tres tipos de colegio, a un nivel de significación de $0'05$.

6.2. Estudie las diferencias en el índice de analfabetismo de cuatro pueblos de la sierra, sabiendo que, tras elegir aleatoriamente una muestra de cada uno de ellos y someterlos a un test, los resultados fueron:

Pueblo 1	Pueblo 2	Pueblo 3	Pueblo 4
78	52	82	57
85	48	91	61
90	60	85	45
77	35	74	46
69	51	70	
	47		

6.3. Se quiere realizar un experimento para averiguar las producciones de cuatro variedades de cebada diferentes: A, B, C y D. Para ello, se asigna aleatoriamente cada una de las variedades a varias parcelas igualmente fértiles. Determine si existen diferencias en las producciones de cebada y si alguna destaca más que otra.

A	B	C	D
35	41	41	38
28	25	39	35
37	42	52	27
29	33	61	31
36	37	63	30
34	40	57	35
		55	29

6.4. Un padre quiere saber si a su hijo se le dan bien todas las asignaturas o hay alguna que se le de peor, para apuntarlo en clases particulares el año siguiente. Las notas de este curso son:

Fís. y Quím.	Biología	Matemát.	Inglés
72	81	88	74
80	74	82	71
83	77	90	77
75		87	70
		80	

Estudie si existen diferencias entre las notas de las distintas asignaturas a un nivel de significación del 1 %.

6.5. Una fábrica quiere averiguar si el tiempo de vida de los tres tipos de bombillas que producen son iguales o por el contrario existen diferencias significativas. Para ello, toman muestras aleatorias de las vidas medias (en horas) de cada una de ellas, obteniendo:

Tipo I	407	411	409		
Tipo II	404	406	408	405	402
Tipo III	410	408	406	408	

Estudie, a nivel de significación $0'05$, si existen tales diferencias.

6.6. Una fábrica produce 4 modelos diferentes de radios con las mismas prestaciones. Debido a problemas económicos se ve obligado el año que viene a producir sólo tres, por lo que está interesado en realizar un estudio sobre las preferencias de los consumidores. Tras seleccionar aleatoriamente algunos consumidores y pedirles que puntuaran del 1 al 9 según les gustara el modelo, los resultados fueron los siguientes:

Modelo 1	Modelo 2	Modelo 3	Modelo 4
8	7	8	6
8	8	6	5
9	7	3	1
7	7	4	2
9	6	5	1
	8	4	3
		5	

Estudie si los cuatro modelos gustan por igual a los consumidores.

6.7. Se está realizando un estudio sobre el tiempo que tarda una determinada proteína en ser asimilada. Se han establecido tres niveles de dicha proteína: 5, 10 y 15 miligramos. Tras seleccionar una muestra de 12 individuos e inyectarles aleatoriamente la proteína, en sus distintos niveles, se midió el tiempo de asimilación (en minutos), obteniendo:

Proteína		
5	10	15
3	5	4
5	6	7
7	8	5
6	7	9

Suponiendo que los tiempos de asimilación se distribuyen normalmente:

a) Compruebe si se verifica la condición de homocedasticidad.

b) ¿Son las observaciones independientes?

c) Estudie si los efectos de las distintas dosis son los mismos.

6.8. Se quiere realizar un experimento para mejorar el rendimiento de los jugadores de baloncesto. Dicho experimento consiste en someter a los jugadores a tres tipos de técnicas de relajación antes de comenzar los partidos y medir el rendimiento por el número de canastas metidas en ese partido. Tras elegir 12 jugadores y someterlos aleatoriamente a cada una de las técnicas, el número de balones encestados por cada uno fue el siguiente:

Técnicas		
A	B	C
19	24	17
23	27	28
30	29	31
32	30	34

Averiguar si las técnicas influyen de igual modo en el rendimiento de los jugadores.

6.9. Una conocida empresa piensa sacar a la venta un nuevo producto y tiene pensado hacer propaganda por tres medios de comunicación: prensa, radio y televisión. Le interesa saber cuál de todas las propagandas resulta más atractiva, por lo que selecciona aleatoriamente a 5 personas y les muestra cada una de las propagandas, pidiéndoles que puntúen de 1 a 10 según les haya gustado. Los resultados fueron los siguientes:

Televisión	Prensa	Radio
$9'3$	$4'8$	7
9	7	$5'7$
$9'7$	$7'3$	4
$7'7$	$6'8$	$4'8$
$8'8$	$6'5$	5

Estudie si las tres propagandas son igualmente interesantes.

6.10. Un grupo de psicólogos está estudiando tres métodos para curar la timidez, los métodos A, B y C. Para ello seleccionan tres muestras de pacientes correspondientes a los tres métodos, obteniendo los resultados siguientes:

Método A	Método B	Método C
$9'5$	$5'5$	7
$6'5$	$4'5$	$3'75$
8	$7'25$	$5'5$
$7'5$	$7'75$	8
6	9	$2'5$
$4'5$	6	3
7	8	5
$8'5$	$4'25$	$3'25$
8	7	$4'5$
	$8'75$	$6'5$
	$6'25$	

a) Compruebe si las observaciones son independientes y Normales.

b) Estudie si los tres métodos son igualmente eficaces.

c) En caso negativo, estudie entre qué métodos existen diferencias significativas.