



CEU

*Universidad
Cardenal Herrera*

***Introducción a la Inferencia
estadística
(intervalos de confianza y p-
valor).***

ÍNDICE

	Página
1. Introducción a la inferencia estadística	3
a. Población y muestra	3
b. Intervalos de confianza.....	6
c. Contrastes de hipótesis.....	8
2. Comparación de dos o más poblaciones	12
a. Comparación de dos proporciones.....	12
b. Test Chi-cuadrado.....	14
c. Comparación de varianzas. Test de Levene.....	16
d. Test t de comparación de dos medias.....	18
e. ANOVA.....	25
3. Tareas recomendadas	29



CEU

Universidad
Cardenal Herrera

(Resumen y aplicación con R-Commander)

1. Introducción a la Inferencia Estadística

a. Población y muestra.

Llamamos **población estadística**, *universo* o *colectivo* al conjunto de referencia del que extraemos las observaciones, es decir, el conjunto de todas las posibles unidades experimentales. Por más que nos refiramos muchas veces a este concepto como población, este conjunto no tiene que ser necesariamente un grupo de personas o animales (pensemos en las variables *Cantidad de plomo en orina*, *Procedimiento quirúrgico*, *Visitas al médico*, *Tiempo hasta que muere una persona tras una operación*).

Llamamos **muestra** a un subconjunto de elementos de la población que habitualmente utilizaremos para realizar un estudio estadístico. Se suelen tomar muestras cuando es difícil, imposible o costosa la observación de todos los elementos de la población estadística, es decir, su uso se debe a que frecuentemente la población es demasiado extensa para trabajar con ella. El número de elementos que componen la muestra es a lo que llamamos *tamaño muestral* y se suele representar por la letra minúscula n .

Nuestro propósito será llegar a conocer ciertas características de la población a partir de la muestra que dispongamos. A este proceso le llamamos **inferencia**.

Ejemplo

Estudio de enfermos renales

Si quisiéramos conocer las características de los enfermos renales en cuanto a *calidad de vida*, *tipo de tratamiento*, *edad de aparición de la enfermedad*, *sexo*, *variables que influyen en el éxito de un trasplante*,..., difícilmente podríamos acceder a todos y cada uno de los enfermos renales que existen (sería la *población* en estudio), pero posiblemente podríamos conseguir a través de algunos hospitales o centros de hemodiálisis los datos de una cantidad determinada de este tipo de enfermos (por ejemplo $n = 200$ enfermos). Nuestro objetivo no sería conocer las características de esos 200 enfermos en concreto, pero utilizaríamos el conocimiento sobre estos 200 enfermos para *obtener conclusiones* sobre todos los enfermos renales



CEU

Universidad
Cardenal Herrera

(nuestra *población* a estudio). Este proceso es lo que se conoce como *inferencia estadística*.

Estadísticos, estimadores y parámetros.

Un **estadístico** es una medida usada para describir alguna característica de una muestra (media, mediana, desviación típica,...) y un **parámetro** es una medida usada para describir las mismas características pero de la población (media, mediana, desviación típica,...). Cuando el estadístico se calcula en una muestra con idea de hacer inferencia sobre la misma característica en la población, se le llama **estimador**. La inferencia estadística pretende aproximarse a los parámetros de la población a partir de los estimadores de la muestra. Para distinguir los estimadores (valores muestrales) de los parámetros (valores poblacionales) los representaremos a partir de ahora con diferentes símbolos:

Característica	Muestra (Estadístico)	Población (Parámetro)
Variable Cuantitativa		
Media	\bar{x}	μ
Desviación típica	s	σ
Varianza	s^2	σ^2
Variable Categórica		
Porcentaje	\hat{P}	P

Variación entre muestras.

Si tomamos varias muestras aleatorias de cierta población, cualquier estimador tomará valores distintos para cada una de ellas. A esta variación en las estimaciones, efecto del azar sobre la elección de la muestra, se le llama **variación muestral**. La variación muestral dependerá de la variabilidad de la variable que tengamos y también del tamaño de la muestra.



CEU

Universidad
Cardenal Herrera

Ejemplo

Ejemplos de variación entre muestras

Si tomamos distintas muestras de la temperatura corporal en población sana tendremos una variación muestral bastante baja (la variabilidad de esta variable es baja entre diferentes personas).

Si tomamos la tensión arterial en la población española obtendremos una variación muestral bastante más elevada.

(Variabilidad de esta variable alta).

Si tomamos muestras de tamaño 10 y calculamos medias muestrales, por ejemplo, se parecerán entre ellas menos, que si las muestras que tomamos son de tamaño 1000. Es decir la variación muestral sería en general más baja cuanto más grande sea la muestra utilizada para calcular nuestro estimador.

Distribución de estadísticos en el muestreo.

▣ Error estándar de la media muestral

El *Teorema Central del Límite* nos asegura que si nuestra muestra es *razonablemente grande* la distribución de la media muestral de cualquier variable sigue una distribución *Normal* y que además, la desviación típica de esta media tiene como expresión:

$$\frac{\sigma}{\sqrt{n}}$$

que representa la desviación típica de la variable original y n es el tamaño de la muestra. A la expresión anterior se le llama *error estándar* de la media.

Supongamos que tenemos una variable cuantitativa cualquiera X , cuya media en la población es μ y cuya desviación típica (también en la población) es σ . Si se toman varias muestras de tamaño *suficientemente grande* y llamamos \bar{X} a la variable que guarda las medias muestrales para cada una de las muestras, por el Teorema Central del Límite tenemos asegurado:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

▣ Error estándar de un porcentaje

En el caso de que la variable de interés sea una variable nominal no tiene sentido que nos planteemos el error estándar de su media (de hecho la media de una variable nominal no tiene tampoco sentido) sino el de su



CEU

Universidad
Cardenal Herrera

porcentaje de individuos en cada uno de sus valores. En este caso si P es el porcentaje de respuestas en ese valor su error estándar será:

$$\sqrt{\frac{P \cdot (100 - P)}{n}}$$

En la expresión anterior se ha supuesto que la variable P está expresada en tantos por 100, si estuviera expresada en tantos por uno (es decir P es un valor entre 0 y 1) únicamente habríamos de cambiar en ella el valor 100 por 1 y la expresión seguiría siendo válida.

Supongamos que tenemos una variable categórica y que nos interesa estimar el porcentaje de una de sus categorías en la población, al que llamamos P . Si tomamos varias muestras de tamaño *suficientemente grande* (n) y en cada una de esas muestras obtenemos una estimación del porcentaje de interés, si llamamos \hat{P} a la variable que guarda los porcentajes de esas muestras, se cumple que esta variable aleatoria sigue la siguiente distribución:

$$\hat{P} \sim N\left(P, \sqrt{\frac{P \cdot (100 - P)}{n}}\right)$$

b. Intervalos de confianza

El proceso de inferencia es aquel mediante el cual se pretende estimar el valor de un parámetro a partir del valor de un estadístico. Esta estimación puede ser puntual o bien por intervalo. La mejor **estimación puntual** de un parámetro es simplemente el valor del estadístico correspondiente, pero es poco informativa porque la probabilidad de no dar con el valor correcto es muy elevada, es por eso que se acostumbra a dar una estimación por intervalo, en el que se espera encontrar el valor del parámetro con una elevada probabilidad. Esta estimación recibe el nombre de estimación mediante **intervalos de confianza**.

La estimación por **intervalos de confianza** consiste en determinar un posible rango de valores o intervalo ($a; b$), en el que, con una determinada probabilidad, sus límites contendrán el valor del parámetro poblacional que andamos buscando. Para cada muestra obtendremos un intervalo distinto que, para el X % de ellas, contendrá el verdadero valor del parámetro. A este intervalo se le denomina **intervalo de confianza**.

Evidentemente esta técnica no tiene porqué dar siempre un resultado correcto, tal y como hemos comentado para algunas muestras el intervalo correspondiente contendrá el verdadero valor del parámetro y para otras no. A la probabilidad de que hayamos acertado al decir que el intervalo



CEU

Universidad
Cardenal Herrera

contiene al parámetro se la denomina **nivel de confianza** (o simplemente **confianza**). También se denomina **nivel de significación** a la probabilidad de errar en esta afirmación, es decir la significación (probabilidad de errar con nuestro intervalo) será igual a 1-(nivel de confianza), ya que el nivel de confianza corresponde a la probabilidad de que el intervalo contenga el valor verdadero del parámetro.

Ejemplo.

En un estudio se pretende estimar la edad media a la que se diagnostica la Diabetes Mellitus en la Comunitat Valenciana. Para ello se dispone de una muestra de 21 pacientes a los que se les ha preguntado la edad de diagnóstico de la enfermedad. A partir de estos 21 pacientes se ha obtenido una edad media (muestral) de 48.78 años y una desviación típica muestral de 16.32. Calcula un intervalo de confianza al 95% para la edad media de diagnóstico de esta enfermedad en la región de estudio.

(Mediante la distribución t-Student en este caso hallamos el intervalo;)

[41.35 , 56.20]

Con un 95% de confianza, la edad media a la que se diagnostica la Diabetes Mellitus en la Comunitat Valenciana estará contenida en el intervalo [41.35, 56.20], es decir, entre 41 y 56 años aproximadamente.

Ejemplo.

Un estudio pretende estimar el porcentaje de hipertensos que hay entre las personas mayores de 65 años en la Comunidad Valenciana. Además de una estimación puntual de este porcentaje, interesa obtener un intervalo de confianza al 95% para este parámetro de la población (P). Para llevar a cabo este estudio, han sido seleccionadas 350 personas mayores de 65 años en toda la Comunidad, resultando tras realizar las pruebas correspondientes que 167 padecen de hipertensión.

Estimador puntual: $(167/350) \times 100 = 47.71\%$

*Intervalo de confianza: **[42.48, 52.94]***

Con un 95% de confianza, el porcentaje de hipertensos entre las personas mayores de 65 años en la Comunidad Valenciana estaría contenido en el intervalo [42.48, 52.94], es decir, aproximadamente entre el 42.5% y 53% de la población.



CEU

Universidad
Cardenal Herrera

c. **Contrastes de hipótesis**

Los contrastes de hipótesis son capaces de responder a preguntas concretas que nos podemos formular sobre los parámetros poblacionales de interés, por ejemplo: *¿La cantidad media diaria de sal ingerida por hipertensos es mayor que la que ingieren las personas con presión arterial normal?, ¿La temperatura corporal de los pacientes que han sufrido cierta infección bacteriana es superior a los 36.7 grados centígrados?, ¿La proporción de personas diabéticas con problemas de vista es superior a la de la población general?*. Resulta evidente que un mecanismo capaz de dar respuesta a cuestiones como las anteriores sería una herramienta muy valiosa, en consecuencia los contrastes o tests de hipótesis son una de las utilidades más valoradas y extendidas en la realización de estudios estadísticos.

Elementos fundamentales de un contraste de hipótesis:

▣ Las hipótesis

En cualquier contraste de hipótesis tendremos 2 alternativas complementarias en las que se especificarán distintos valores de un parámetro poblacional y a la vista de los datos habremos de optar por una de ellas. Por ejemplo, si deseamos conocer si el valor de un parámetro μ puede ser igual a 25 o por el contrario es inadmisibile a la vista de los datos que disponemos, nuestras hipótesis serán:

$$\mu=25 \text{ y } \mu \neq 25$$

Estas 2 hipótesis que hemos señalado no jugarán el mismo papel dentro de cualquier contraste de hipótesis, y por tanto cada una de ellas recibirá un nombre específico:

Hipótesis nula, a la que habitualmente nos referimos como ***H0***.

Hipótesis alternativa, a la que habitualmente nos referimos como ***HA*** o ***H1***.

A la hipótesis nula siempre se le concederá el beneficio de la duda e intentaremos encontrar en nuestra muestra evidencias en contra de ella. Así, al terminar el contraste habremos de optar por aceptar *H0* (si no tenemos evidencia suficiente en su contra) o rechazarla (si los datos hacen que la descartemos).

Se podría hacer un símil entre el papel de la hipótesis nula en un contraste de hipótesis y el acusado de un juicio: ambos tienen presunción de inocencia y si los datos no aportan evidencias suficientes en contra de su veracidad nos veremos obligados a aceptarlos. En consecuencia, si en un contraste de hipótesis rechazamos la hipótesis nula será porque disponemos de evidencias suficientes en su contra, es decir estamos razonablemente seguros de que dicha hipótesis es falsa. Por el contrario si aceptamos *H0*



CEU

Universidad
Cardenal Herrera

será porque no hemos encontrado evidencias suficientes en su contra, pero esto no implica que estemos más o menos seguros de que realmente dicha hipótesis sea cierta, podría darse el caso de que H_0 fuera falsa pero que los datos no aportan evidencia suficiente como para que lleguemos a dicha conclusión. En los juicios también pasa algo parecido con los acusados, si alguien resulta absuelto en un juicio no será porque hemos determinado su inocencia sino porque no hemos encontrado pruebas suficientes que lo inculpen. En el siguiente cuadro se resumen las conclusiones a que conduce cada posible resultado de un contraste de hipótesis:

Resultado del contraste	Conclusión
Rechazar H_0	Podemos descartar H_0
Aceptar H_0	Aceptamos la posibilidad de que H_0 sea cierta aunque también lo podría ser H_1

Tal y como se puede apreciar en el cuadro anterior el rechazar H_0 conduce a conclusiones mucho más valiosas que el aceptarlo. Cuando aceptamos H_0 seguimos sin saber cual de las dos opciones, la hipótesis nula o la alternativa, admitimos como cierta; por el contrario, cuando rechazamos H_0 estamos admitiendo implícitamente como cierta H_1 , de esta forma nos decantamos por una de las dos hipótesis. Por este motivo suele ser bastante más valorado un resultado en el que se rechaza la hipótesis nula que aquel en el que se acepta, es decir el objetivo habitual que se perseguirá a la hora de hacer cualquier contraste de hipótesis será el intentar descartar la hipótesis nula que nos planteemos.

▣ La unilateralidad o bilateralidad del contraste

Tal y como hemos podido comprobar hasta ahora todas las hipótesis que hemos formulado han sido expresadas 'matemáticamente' como relaciones de igualdad o desigualdad entre un parámetro y un valor concreto. Como norma general, y por razones que justificaremos con mayor detalle en la próxima sección, la hipótesis nula se corresponderá siempre con una igualdad. Sin embargo, la hipótesis alternativa no ha de responder siempre a una relación de desigualdad completa (\neq) sino que puede responder simplemente a una desigualdad parcial ($<$ o $>$). El utilizar una u otra desigualdad dependerá del problema en particular, en concreto de aquello que queramos demostrar.

Aquellos contrastes en los que la hipótesis alternativa se defina mediante el signo \neq se llaman **Contrastes bilaterales**, ya que nos valen ambos sentidos de la desigualdad (tanto si el primer término es mayor que el segundo, o menor). Por el contrario aquellos contrastes en los que la hipótesis nula sea de la forma $<$ o $>$ se conocen como **Contrastes unilaterales**.

▣ La significatividad



CEU

Universidad
Cardenal Herrera

Según hemos comentado previamente el objetivo fundamental de los contrastes de hipótesis será cuantificar la fiabilidad con la que podemos aceptar la hipótesis nula. Dicha fiabilidad, según veremos en la próxima sección, se mide como la probabilidad que tendríamos de equivocarnos en nuestra decisión si rechazáramos la hipótesis nula.

Obviamente cuando dicha probabilidad sea 'alta' no rechazaremos H_0 ya que tendríamos un gran riesgo de equivocarnos. Por el contrario si la probabilidad de errar en caso de rechazar H_0 fuera muy 'baja' podríamos rechazarla sin temor. Esta es la idea fundamental de los contrastes de hipótesis.

En cualquier caso para llevar a cabo el procedimiento anterior hemos de determinar cual será el umbral para la probabilidad por debajo del cual consideraremos que el riesgo de equivocarnos es 'bajo' o no. Dicho valor se conoce como la *significatividad* del contraste y habitualmente se denota como α . La interpretación de este parámetro sería: *Máxima probabilidad de equivocarnos que estamos dispuestos a asumir en caso de que rechazemos la hipótesis nula.*

En la práctica totalidad de estudios estadísticos el valor que se suele elegir para α es 0.05, aunque también suelen tomarse $\alpha = 0.01$ o $\alpha = 0.10$ dependiendo de si queremos asumir menos o más riesgo de equivocarnos, respectivamente, en caso de rechazar la hipótesis nula. La utilización de estos valores se ha definido por consenso de la comunidad científica y resulta muy inusual la utilización de otros valores de significatividad distintos a los anteriores y su utilización requiere la existencia de alguna razón de peso que habría de ser debidamente justificada.

▣ p-valor

En todo contraste de hipótesis aceptaremos o rechazaremos la hipótesis nula dependiendo del valor que hayamos establecido de significatividad. En concreto, si la significatividad es más alta admitimos mayor riesgo de equivocarnos cuando rechazamos la hipótesis nula y en consecuencia rechazaremos dicha hipótesis con mayor facilidad.

El *P-valor* de un contraste de hipótesis se define como la probabilidad de error en que incurriríamos en caso de rechazar la hipótesis nula con los datos de que disponemos. La importancia del P-valor viene dada porque nos proporciona un resultado mucho más informativo que el que nos proporciona el propio resultado del contraste, ya que éste termina diciendo únicamente si aceptamos o no la hipótesis nula ya sea con una gran holgura, o sin ella. Sin embargo el P-valor cuantifica el riesgo a equivocarnos que tendremos que asumir si queremos rechazar H_0 . Por tanto se suele interpretar el P-valor como una medida de la evidencia que aportan los datos a favor de la hipótesis nula, en concreto, aquellos valores bajos del P-valor se corresponden con datos que no apoyan la hipótesis nula, ya que la probabilidad de equivocarnos en caso de que la rechazáramos sería baja.



CEU

*Universidad
Cardenal Herrera*

El P-valor supone además una herramienta alternativa para la resolución de contrastes de hipótesis. Así, supongamos pues que disponemos del valor del P-valor p de cierto contraste y supongamos que dicho valor es inferior a la significatividad del contraste, es decir $p < \alpha$, en ese caso la probabilidad de equivocarnos en caso de rechazar la hipótesis nula (el P-valor) es menor que la probabilidad de equivocarnos que estaríamos dispuestos a asumir en caso de rechazar la hipótesis nula (la significatividad), por tanto podremos rechazar la hipótesis nula. Por el contrario, si el P-valor es mayor que la significatividad la probabilidad de equivocarnos en caso de rechazar la hipótesis nula sería superior al riesgo de equivocarnos que querríamos asumir, por lo que no deberíamos rechazar dicha hipótesis.

$P - \text{valor} < \alpha \Rightarrow \text{Rechazamos } H_0$

$P - \text{valor} > \alpha \Rightarrow \text{Aceptamos } H_0$



CEU

Universidad
Cardenal Herrera

2. Comparación de dos o más poblaciones

a. Comparación de dos proporciones.

La situación en estudio estará compuesta por dos poblaciones independientes, *Población 1* y *Población 2*. El objetivo es comparar los porcentajes de cierta respuesta de una variable cualitativa de interés en cada una de ellas, a los que llamaremos P_1 y P_2 . Para llevar a cabo esta comparación dispondremos de dos muestras, una cada una de las poblaciones en estudio, con tamaños que denotaremos n_1 y n_2 . En cada una de esas muestras podremos obtener el porcentaje de interés: \hat{P}_1 y \hat{P}_2 .

Para llevar a cabo la comparación, podemos tanto plantear un contraste de hipótesis (unilateral o bilateral dependiendo de la situación):

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2 \quad (H_1 : P_1 < P_2 ; H_1 : P_1 > P_2)$$

o bien calcular un intervalo de confianza para la diferencia de ambos porcentajes, es decir, para $P_1 - P_2$.

Ejemplo:

Datos: [ObesidadInfantil2.xls/ObesidadInfantil2.rda](#)

Pregunta: ¿El porcentaje de niños obesos es el mismo entre los que utilizan el servicio de comedor y entre los que no lo utilizan? ($\alpha=0.05$)

Variable 1: Comedor (Define las dos poblaciones: Comedor Sí/Comedor No)

Variable 2: Obeso (Tenemos interés en el % de obesos en cada población)

P_1 =%de niños obesos entre los que utilizan servicio de comedor escolar

P_2 =%de niños obesos entre los que NO utilizan servicio de comedor escolar

H0: $P_1 = P_2$

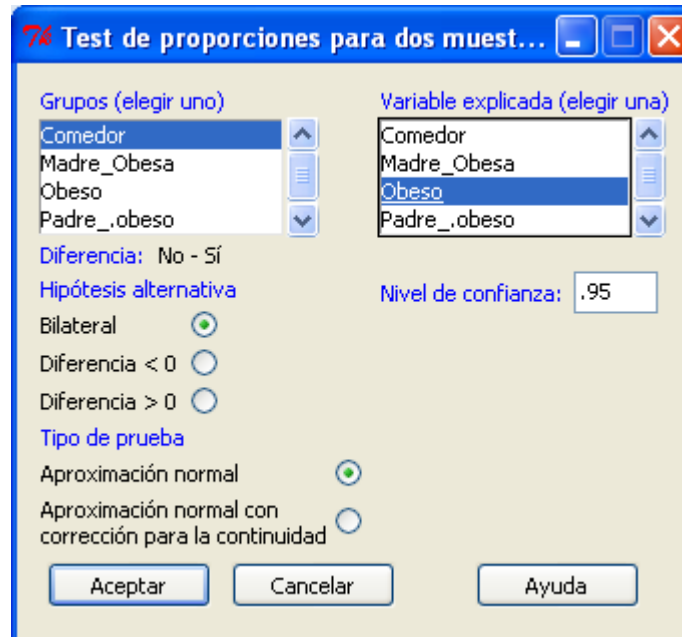
H1: $P_1 \neq P_2$

R-Commander: Estadísticos> Proporciones> Test de proporciones para dos muestras...



CEU

Universidad
Cardenal Herrera



Resultados en R-Commander:

```
Ventana de resultados
> prop.test(.Table, alternative='two.sided', conf.level=.95, correct=FALSE)

      2-sample test for equality of proportions without continuity correction

data: .Table
X-squared = 0.114, df = 1, p-value = 0.7357    p-valor
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.2398826  0.3398826    Intervalo de confianza para P1-P2
sample estimates:
prop 1 prop 2
 0.70  0.65
```

Interpretación de los resultados:

- o p-valor=0.7353

Como $p\text{-valor} > \alpha$, por lo que no podemos rechazar la hipótesis nula. Es decir, no tenemos evidencias suficientes para concluir que existan diferencias significativas entre el porcentaje de niños obesos que utilizan el servicio de comedor y el porcentaje de niños obesos que no lo utilizan.

- o Intervalo de confianza al 95% para P1-P2: (-0.2398,0.3398)

Como el 0 está contenido en este intervalo, no podemos descartar este valor para la diferencia entre P1 y P2, por lo que no podemos concluir que existan diferencias significativas entre el porcentaje de niños obesos que utilizan el servicio de comedor y el porcentaje de niños obesos que no lo utilizan.

b. Test Chi-cuadrado

La prueba Chi-cuadrado (χ^2) permite determinar si dos variables cualitativas (categóricas) con dos o más categorías cada una están o no asociadas.

El contraste de hipótesis que nos planteamos es:

H₀: Las variables son independientes.

H₁: Las variables no son independientes

Ejemplo:

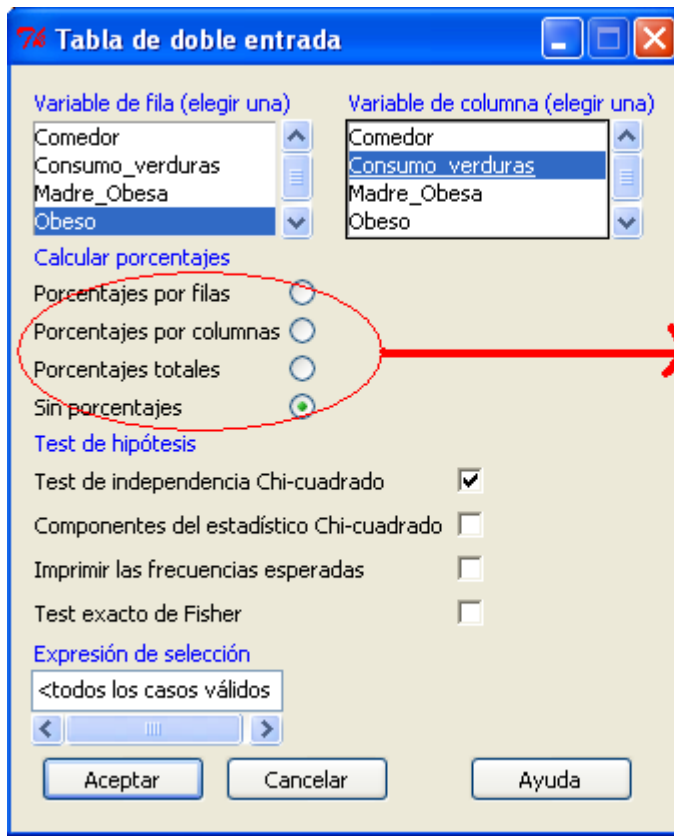
Datos: [ObesidadInfantil2.xls/ObesidadInfantil2.rda](#)

Pregunta: ¿Existe relación entre las variables consumo de verduras y la que indica la obesidad o no de los niños? ($\alpha=0.05$)

H₀: Las variables son independientes

H₁: Las variables no son independientes

R-Commander: Estadísticos > Tablas de contingencia > Tabla de doble entrada...



Seleccionamos la opción deseada

Resultados en R-Comander:

```
Ventana de resultados
> .Table
      consumo_verduras
Obeso Escaso Habitual Nunca
No      8      15      4
Sí      3       1      9

> .Test <- chisq.test(.Table, correct=FALSE)
> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 13.1576, df = 2, p-value = 0.001390
```

Tabla de contingencia

p-valor

Interpretación de los resultados:

- o p-valor=0.001390



CEU

*Universidad
Cardenal Herrera*

Como $p\text{-valor} < \alpha$, rechazamos la hipótesis nula. Por tanto, podemos concluir que existe una relación significativa entre las variables Obesidad y Consumo de verduras.

c. Comparación de dos varianzas.

Supongamos que queremos comparar las varianzas poblacionales (σ_1^2 y σ_2^2) de dos poblaciones normales. Para ello dispondremos de dos muestras, una de cada población, y por tanto tendremos conocimiento de los dos tamaños de ambas muestras (n_1 y n_2) y de las varianzas muestrales obtenidas (S_1^2 y S_2^2).

El contraste de hipótesis que nos plantearemos será el siguiente:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Ejemplo:

Datos: ObesidadInfantil2.xls/ObesidadInfantil2.rda

Pregunta: Suponiendo que el peso al nacer sigue una distribución Normal ¿La varianza en el peso al nacer de los niños nacidos de madres obesas es significativamente diferente de la varianza en el peso al nacer de los niños nacidos de madres no obesas? ($\alpha=0.05$)

σ_1^2 = Varianza del peso al nacer de los niños nacidos de madres obesas

σ_2^2 = Varianza del peso al nacer de los niños nacidos de madres no obesas

Nos planteamos el contraste:

H0: $\sigma_1^2 = \sigma_2^2$

H1: $\sigma_1^2 \neq \sigma_2^2$

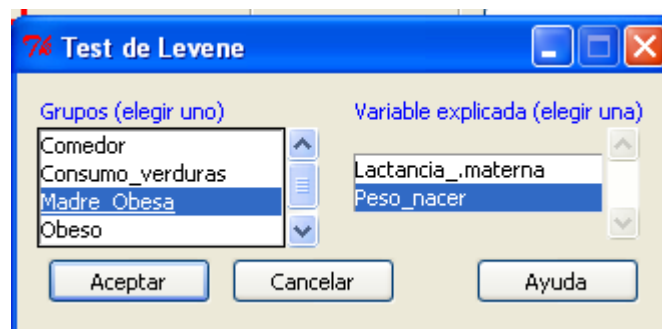
Para comparar dos varianzas en R-Commander tenemos dos opciones:

- R-Commander: Estadísticos > Varianzas > Test F para dos varianzas...

ó

- R-Commander: Estadísticos > Varianzas > Test de Levene...

(El test de Levene se utiliza para la comparación de dos o más varianzas)





CEU

Universidad
Cardenal Herrera

```
Ventana de resultados  
  
> levene.test(ObesidadInfantil2$Peso_nacer, ObesidadInfantil2$Madre_Obesa)  
Levene's Test for Homogeneity of Variance  
  Df F value Pr(>F)  
group 1  0.1955 0.6609 p-valor  
      38
```

Interpretación de los resultados:

- o p-valor=0.6609

Como $p\text{-valor} > \alpha$, no podemos rechazar la hipótesis nula. Por tanto, no podemos concluir que existan diferencias significativas entre las varianzas del peso al nacer de los niños nacidos de madres obesas y no obesas.

Nota.- En general es habitual realizar un contraste sobre igualdad de varianzas, no con el propósito directo de conocer el comportamiento de la variabilidad de una variable en dos o más grupos diferentes, sino para comprobar si esta hipótesis de igualdad de varianzas es asumible y poder aplicar otras pruebas que necesitan de la misma.

d. *Test t de comparación de dos medias.*

MUESTRAS INDEPENDIENTES

Uno de los análisis estadísticos más comunes en la práctica es probablemente el utilizado para comparar dos grupos independientes de observaciones con respecto a una variable numérica.

Así, si queremos comparar dos medias poblacionales μ_1 y μ_2 , plantearemos el contraste de hipótesis:

H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$

La aplicación de un contraste paramétrico de este tipo requiere **la normalidad de las observaciones para cada uno de los grupos**. La comprobación de esta hipótesis puede realizarse tanto por métodos gráficos (por medio de histogramas, diagramas de cajas o gráficos de normalidad) como mediante tests estadísticos (test de Kolmogorov-Smirnov, test de Shapiro-Wilks). Un número suficiente de observaciones (digamos mayor de 30) justifica la utilización del mismo test. Así mismo, este tipo de metodología exigirá que **la varianza en ambos grupos de observaciones sea la misma**. En primer lugar se desarrollará el test t de Student para el caso en el que se verifiquen ambas condiciones, discutiendo posteriormente el modo de abordar formalmente el caso en el que las varianzas no sean similares.

Cuando no es posible asumir que las varianzas de la variable cuantitativa en las dos poblaciones a comparar son iguales, es recomendable buscar una transformación de la misma que haga posible asumir esta hipótesis (de igualdad de varianzas) como cierta. En caso de no lograr hallar una transformación que haga posible esto, es mejor recurrir a otra metodología (**métodos no paramétricos** como el test de Wilcoxon, que por motivos de tiempo no son objeto del temario de esta materia) para realizar la comparación de las medias.

Ejemplo:

Datos: [ObesidadInfantil2.xls/ObesidadInfantil2.rda](#)

Pregunta: ¿Existen diferencias significativas en el número medio de meses de lactancia entre los niños que en la actualidad son obesos y los que no lo son? ($\alpha=0.05$)



CEU

Universidad
Cardenal Herrera

Se trata de una comparación de medias de muestras independientes puesto que los niños obesos y los no obesos son grupos poblacionales diferentes y no guardan relación entre sí.

En primer lugar comprobaremos si se cumplen las hipótesis de aplicación de un test t de comparación de muestras independientes: Normalidad e Igualdad de Varianzas (a esta última hipótesis también se le suele llamar "Hipótesis de Homocedasticidad" o "Hipótesis de Homogeneidad de varianzas").

Normalidad: R-Commander: Estadísticos> Resúmenes> Test de Normalidad de Shapiro-Wilks...

Para estudiar si el número de meses de lactancia se distribuye según una distribución normal en cada una de las dos poblaciones (niños obesos y no obesos), aplicamos, por ejemplo, la prueba Shapiro - Wilks. R Commander tiene la posibilidad de seleccionar la prueba para una variable numérica, mediante la ruta **Estadísticos / Resúmenes / Test de normalidad Shapiro – Wilks**, pero no de seleccionar los distintos grupos que determina el factor (niño obeso o no), con lo que es necesario escribir directamente las sentencias necesarias en la ventana de instrucciones y ejecutarlas.

El test de Shapiro-Wilks se plantea el contraste:

H0: Los datos siguen una distribución Normal

H1: Los datos no siguen una distribución Normal

Si seleccionamos con la variable "Lactancia_materna" la opción de R-Commander para el test de Shapiro-Wilks aparecerá:



The screenshot shows the R Commander interface. The menu bar includes 'Fichero', 'Editar', 'Datos', 'Estadísticos', 'Gráficas', 'Modelos', 'Distribuciones', and 'Herramientas'. The 'Conjunto de datos' field is set to 'ObesidadInfantil2'. The 'Ventana de instrucciones' (Instructions window) contains the following code:

```
load("C:/Documents and Settings/Administrator/My Docume:  
shapiro.test(ObesidadInfantil2$Lactancia_.materna)
```

The second line of code is circled in red. To its right, the text 'Normalidad de la variable Lactancia Materna' is written in blue. Below the instructions window is the 'Ventana de resultados' (Results window), which displays the following output:

```
> load("C:/Documents and Settings/Administrator/My Docu  
> shapiro.test(ObesidadInfantil2$Lactancia_.materna)  
  
Shapiro-Wilk normality test  
  
data: ObesidadInfantil2$Lactancia_.materna Resultado  
W = 0.9334, p-value = 0.02085
```

Pero lo que necesitamos poder asumir es que la variable "Lactancia_materna" se comporte de forma "Normal" en cada grupo, es decir, en el grupo de "Obesos" y en el de "No obesos". Para ello, añadimos la siguiente expresión en la *Ventana de instrucciones* y pulsamos ejecutar:



```
Ventana de instrucciones
load("C:/Documents and Settings/Administrator/My Documents/My Dropbox/MasterPaloma/
shapiro.test(ObesidadInfantil2$Lactancia_.materna)
shapiro.test(ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso=="Si"])
shapiro.test(ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso=="No"])

Ventana de resultados

data: ObesidadInfantil2$Lactancia_.materna
W = 0.9334, p-value = 0.02085

> shapiro.test(ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso=="Si"])

      Shapiro-Wilk normality test

data: ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso ==      "Si"]
W = 0.9584, p-value = 0.7283

> shapiro.test(ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso=="No"])

      Shapiro-Wilk normality test

data: ObesidadInfantil2$Lactancia_.materna[ObesidadInfantil2$Obeso ==      "No"]
W = 0.949, p-value = 0.2023
```

Los resultados indican que no se puede rechazar la Normalidad de la variable "Lactancia materna" ni en el grupo de niños Obesos (p-valor=0.7283), ni en el grupo de niños No Obesos (p-valor=0.2023). Por tanto, es asumible la Normalidad de la variable "Lactancia materna" en ambos grupos (niños obesos y no obesos).

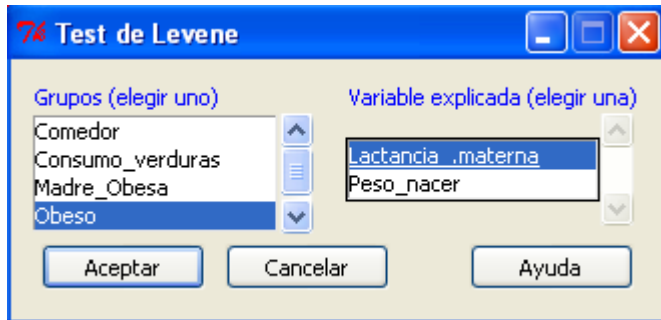
Homogeneidad de varianzas : R-Commander: Estadísticos> Varianzas> Test de Levene...

A continuación comprobaremos si podemos asumir que la variable "Lactancia materna" tiene la misma varianza en los dos grupos que pretendemos comparar.

El contraste que se plantea el test de Levene es:

H0: La varianza de la variable "Lactancia materna" es igual en niños obesos y no obesos

H1: La varianza de la variable "Lactancia materna" es distinta en niños obesos y no obesos



Y tras pulsar "Aceptar":

```
ventana de resultados
> levene.test(ObesidadInfantil2$Lactancia_.materna, ObesidadInfantil2$Obeso)
Levene's Test for Homogeneity of Variance
  Df F value Pr(>F)
group 1 9.411 0.003964 **
  38
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como el p-valor obtenido es 0.003964, menor que 0.05, rechazamos la igualdad de varianzas en ambos grupos y por tanto no podríamos asumir la igualdad de varianzas que necesitamos para aplicar esta prueba. Sin embargo, el test t para muestras independientes en R-Commander incorpora la posibilidad de que la hipótesis de igualdad de varianzas no pueda ser asumida y ya está preparado para poder realizar este test incluso en ese caso.

Comparación de la media poblacional de las dos poblaciones : R-Commander: Estadísticos> Medias> Test t para muestras independientes...

Una vez comprobadas las hipótesis comparamos, por fin, las dos medias mediante el test t para muestras independientes:

- H0: El numero medio de meses de lactancia es igual en niños obesos y no obesos**
- H1: El número medio de meses de lactancia es distinto en niños obesos y no obesos**

Si vamos a la opción correspondiente en R-Commander:



Test t para muestras independientes

Grupos (elegir uno): Comedor, Madre_Obesa, **Obeso**, Padre_obeso

Variable explicada (elegir una): Lactancia .materna, Peso_nacer

Diferencia: No - Sí

Hipótesis alternativa: Bilateral Diferencia < 0 Diferencia > 0

Nivel de confianza: .95

¿Suponer varianzas iguales? Sí **No**

Aceptar Cancelar Ayuda

No podemos asumir que las varianzas son iguales, pues lo hemos comprobado con el test de Levene

Pulsamos "Aceptar" y obtenemos:

```
Ventana de resultados

> t.test(Lactancia_.materna~Obeso, alternative='two.sided', conf.level

Welch Two Sample t-test

data: Lactancia_.materna by Obeso
t = 7.815, df = 34.806, p-value = 3.638e-09 p-valor
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
13.37801 22.77013 Intervalo de confianza al 95% para la diferencia de medias
sample estimates:
mean in group No mean in group Si
23.07407 5.00000

Medias para el grupo de Obesos y No Obesos
de la variable "Lactancia materna"
```

Podemos observar que el p-valor es 0.000..., y por tanto inferior a 0.05, por lo que podemos rechazar que la lactancia materna media del grupo de obesos y no obesos sea igual (rechazamos H₀). Así, tenemos evidencias suficientes para concluir que existen diferencias significativas entre la lactancia materna media de niños obesos y no obesos.

El intervalo para la diferencia de medias (No Obesos-Obesos) es (13.37, 22.77). Los dos extremos "positivos" indican que la diferencia es un valor positivo con un 95% de confianza, por lo que los niños no obesos han obtenido una lactancia materna media superior a los niños obesos.

MUESTRAS DEPENDIENTES O PAREADAS

Diremos que 2 muestras son pareadas si existe alguna relación entre los elementos de ambas muestras que pudiera establecer dependencia entre los valores obtenidos de la variable de estudio. Por ejemplo, si queremos evaluar los efectos de una dieta sobre el peso corporal en cierta población tomaremos el peso a un conjunto de individuos antes de someterlos a dieta. Tras el periodo de dieta pesamos nuevamente a los integrantes del estudio obteniendo así una segunda medición del peso en cada individuo. Así obtenemos 2 muestras de pesos de la población, pero estas 2 muestras tienen una peculiaridad y es que los individuos que las componen están relacionados, es más son los mismos individuos. En este caso diremos que las muestras están pareadas. Para este tipo de problemas en lugar de plantearnos un contraste habitual sobre la igualdad de medias como el que acabamos de comentar en el apartado anterior, restaríamos las 2 mediciones efectuadas a cada persona (o cada par de mediciones relacionadas), de esta forma obtendremos una única muestra de diferencias y contrastaremos si la media de estas diferencias es distinta de 0 o no. Así conseguimos que las observaciones de la variable sean independientes entre sí, reduciendo así cualquier efecto que pudiera tener esta dependencia sobre los resultados del estudio.

La opción en R-Commander para obtener un análisis del tipo de comparación de medias de muestras dependientes o pareadas es:

Estadísticos>Medias>Test t para datos relacionados...



CEU

Universidad
Cardenal Herrera

e. ANOVA (Análisis de la Varianza)

Supongamos que tenemos una variable cuantitativa Y cuyo valor medio queremos comparar en diferentes grupos (definidos por las k categorías de una variable categórica llamada *factor*). El contraste de hipótesis que nos planteamos es:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ para algún } i, j$$

Si el p-valor obtenido de este contraste es menor que α , se rechazaría la hipótesis nula y se concluiría que *al menos dos de las medias difieren entre sí*. Posteriormente, si resulta de interés, habría que valorar cuál o cuáles son las medias entre las que hay diferencias.

Hipótesis necesarias para la aplicación del ANOVA

Muestreo aleatorio

Todos los individuos que componen las observaciones de cada uno de los grupos deben haber sido elegidos de la población y asignados aleatoriamente a cada uno de ellos.

Normalidad

Los valores de la variable se distribuyen *normalmente* (o *siguen una distribución Normal*) en cada uno de los grupos definidos por el factor, es decir, para cada grupo i , $Y_{ij} \sim N(\mu_i, \sigma^2)$. La violación de este supuesto no afecta mucho a las conclusiones del análisis de la varianza si el tamaño de las muestras de cada grupo es relativamente grande (por ejemplo más de 30 datos por grupo). Si el número de datos por grupo no es excesivamente grande es recomendable realizar un contraste de hipótesis como el siguiente:

H_0 : Los valores siguen una distribución Normal

H_1 : Los valores no siguen una distribución Normal

y en caso de no rechazar la hipótesis nula podríamos decir que *no es un disparate* asumir que el comportamiento de los datos es *normal*. Existen diferentes pruebas para contrastar la normalidad, pero una muy extendida en uso es la llamada *Prueba de Shapiro-Wilks*. Si el p-valor obtenido en esta prueba es superior al nivel de significatividad (normalmente $\alpha = 0,05$) no podemos rechazar la hipótesis de normalidad y por tanto podemos asumirla como cierta.



Homocedasticidad

La *Homocedasticidad*, o lo que es lo mismo, la *Homogeneidad de varianzas*, asume que las varianzas de todos los grupos a comparar son homogéneas (es decir, que no se detectan diferencias significativas entre las varianzas de los grupos a comparar). La violación de esta hipótesis impide asumir como correctos los resultados que de este análisis se deriven, y por tanto impiden la utilización de ANOVA. En los casos en los que no se pueda asumir esta premisa como cierta se recomienda utilizar otros métodos de comparación de medias, como los llamados *no paramétricos*, entre los que se encuentra la prueba de *Kruskall-Wallis* y que no forma parte del temario de esta asignatura. Para poder asumir como cierta esta hipótesis, es recomendable realizar el siguiente contraste de hipótesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i, j$$

Existen diferentes pruebas para llevar a cabo este contraste, pero las más extendidas en uso son el *Test de Barlett* y el *Test de Levene*. Si al realizar alguna de estas dos pruebas obtenemos un p-valor superior al nivel de significatividad (normalmente $\alpha = 0,05$) no podremos rechazar la hipótesis de homogeneidad de varianzas y por tanto podemos asumirla como cierta.

Ejemplo

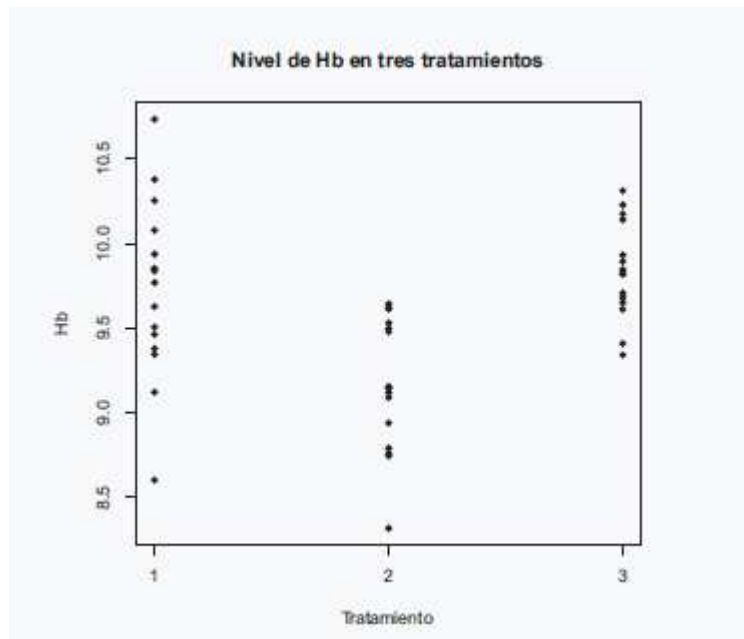
Supongamos que estamos interesados en comprobar si existen diferencias significativas en el nivel medio de hemoglobina (Hb) en tres tratamientos diferentes para personas con cierto tipo de anemia diagnosticada. Con el fin de realizar la comparación correspondiente se toman 45 pacientes con este tipo de anemia diagnosticada y se reparten al azar entre los tres tratamientos (15 en cada grupo). A continuación mostramos el análisis correspondiente para llevar a cabo esta comparación.

En primer lugar mostramos los datos obtenidos del nivel de Hb. en cada uno de los tres grupos:



CEU

Universidad
Cardenal Herrera



Como primera apreciación tras la representación gráfica, sí que parece que el tratamiento 2 obtenga unos niveles de Hb inferiores, de media, a los de los otros dos tratamientos. El análisis adecuado para comprobar si existen diferencias significativas entre los niveles medios de Hb. en los tres tratamientos es un ANOVA, pero en primer lugar debemos comprobar que se cumplen las hipótesis de aplicabilidad de esta técnica:

- Muestreo aleatorio: Esta hipótesis se cumple, pues los 45 pacientes disponibles son asignados aleatoriamente a cada uno de los tres tratamientos.
- Normalidad: Para comprobar esta hipótesis puedo realizar, por ejemplo, un test de Shapiro-Wilks en la que comprobaría si es asumible esta hipótesis, o por lo contrario, los datos la violan de forma clara. Con este fin, para cada uno de los tres grupos plantearía el siguiente contraste de hipótesis:

H_0 : Los valores siguen una distribución Normal

H_1 : Los valores no siguen una distribución Normal



Supongamos que los p-valores obtenidos para este contraste en cada uno de los grupos fueran:

1. Tratamiento 1: Test de Normalidad Shapiro-Wilk (p-valor=0.7645)
2. Tratamiento 2: Test de Normalidad Shapiro-Wilk (p-valor=0.5438)
3. Tratamiento 3: Test de Normalidad Shapiro-Wilk (p-valor=0.8341)

Como en todos los casos, el p-valor para este contraste es mayor que 0.05, en ninguno de los casos puede ser rechazada la hipótesis de normalidad, y por tanto, podemos asumir que se cumplen (al menos estamos seguros que no se desvían demasiado de cumplirse)

- Homocedasticidad u Homogeneidad de varianzas: En este caso, planteamos el contraste de hipótesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i, j$$

Supongamos que realizamos un Test de Levene y el resultado obtenido es: Test de igualdad de varianzas de Levene (p-valor=0.2859). Puesto que el p-valor es mayor que 0.05, no podemos rechazar la hipótesis nula y por tanto no podemos rechazar que la igualdad de varianzas se cumpla. Así, podemos asumir esta hipótesis como cierta (como en el caso anterior, al menos estamos seguros de que en caso de no ser cierta no se desvía mucho de la misma).

Una vez comprobado que las hipótesis necesarias para poder aplicar el ANOVA se cumplen, ya estamos en disposición de aplicar esta técnica y obtener conclusiones sobre la igualdad o no de los niveles medios de la variable cuantitativa sobre los distintos grupos.

Planteamos el contraste ANOVA:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \text{ (Nivel medio de Hb igual en los tres trat.)}$$

$$H_1 : \mu_i \neq \mu_j \text{ para algún } i, j \text{ (Nivel medio de Hb no igual en los tres trat.)}$$

A partir de los datos obtenemos el estadístico F que tiene un valor $F=11.229$. Bajo la hipótesis nula el estadístico F sigue una distribución F con $3 - 1 = 2$ y $45 - 3 = 42$ grados de libertad, es decir una $F_{(2,42)}$. La región de rechazo (para $\alpha = 0,05$) sería aproximadamente $(3,23, +\infty)$, y por tanto con el valor de nuestro pivote rechazaríamos la hipótesis nula. El p-valor resultante es 0,0001.

Conclusión: Tenemos evidencias para concluir que existen diferencias significativas entre los niveles medios de Hb de los tres tratamientos, y por tanto, que al menos hay dos tratamientos entre los que los niveles medios de Hb difieren.

La opción en R-Commander para obtener un análisis del tipo ANOVA es:

Estadísticos>Medias>ANOVA de un factor...

(Los test de Shapiro-Wilks y Levene se solicitan a R-Commander tal y como se ha comentado en la sección anterior)

Tareas recomendadas

A continuación proponemos una serie de tareas cuyo objetivo es que el alumno practique las técnicas de inferencia estadística comentadas en este resumen con el programa R-Commnader.

Datos: TRASPLANTES RENALES (Datos ficticios): Se dispone de una muestra aleatoria de todos los pacientes a los que se les ha realizado un trasplante renal en un determinado hospital cuyos pacientes provienen de tres Áreas de Salud determinadas

Los datos se encuentran en el fichero: **DatosRenales.rda**

Las variables que se han recogido para cada paciente son las siguientes:

- **Edad:** (EDAD) Indica la edad de los pacientes en el momento del trasplante.
- **Sexo:** (SEXO) Hombre (H) o Mujer (M)
- **Área:** (AREASALUD) Área de salud a la que pertenece el paciente (A, B o C). Suponemos las tres áreas cubren a la misma cantidad de población aproximadamente.
- **Días de hospitalización:** (DIASHOSP) Días que estuvo el paciente hospitalizado tras la intervención.
- **Éxito a un año vista:** (EXITOUNAÑO) Éxito del trasplante renal transcurrido el primer año (Sí o No)
- **Tratamiento previo al trasplante:** (TRATPREV) Tratamiento que recibía el paciente antes de realizar el trasplante (Hemodíalisis [Hemo], Diálisis Peritoneal [Peri] y otros [Otros])

Se solicita plantear las técnicas adecuadas y resolver mediante R-Commander para responder a las siguientes preguntas. Comenta todos los resultados obtenidos. Utiliza el nivel de significatividad 0.05 para todas tus conclusiones.

Cuestión 1: ¿Podemos concluir que el porcentaje de Éxitos tras un año es significativamente diferente entre hombres y mujeres?

Cuestión 2: ¿El tratamiento previo al trasplante depende del área de salud o los tratamientos se emplean por igual en las tres áreas?

Cuestión 3: ¿El tiempo medio de hospitalización tras la intervención difiere entre los que tienen éxito al año y los que no?

Cuestión 4: ¿La edad media de los pacientes trasplantados difiere según el tratamiento previo al trasplante?

Cuestión 5: Cuestión libre. Plantea y responde una pregunta que pueda ser respondida mediante las técnicas vistas en esta sesión