ESTUDIO DE LA INFLUENCIA DE LA ADICIÓN DE ADITIVOS EN PASTAS DE CEMENTO Y MORTEROS

Anexo número 1: Metodología del análisis estadístico utilizado para la interpretación de los datos obtenidos

Todos los experimentos realizados examinan algún tipo de variable. Una variable no es más que algo que utilizamos para designar una cantidad susceptible de tomar distintos valores numéricos dentro de un conjunto de números especificado.

Variables dependientes e independientes

Una variable independiente, a veces llamada variable experimental o de predicción, es una variable que es manipulada en un experimento con el fin de observar el efecto sobre una variable dependiente, a veces llamada variable de resultado. Luego una variable dependiente es simplemente una variable que depende de una variable independiente.

La investigación experimental

En la investigación experimental, el objetivo es manipular una variable independiente y luego examinar el efecto que tiene este cambio en una variable dependiente. Puesto que es posible manipular la variable independiente, la investigación experimental tiene la ventaja de permitir a un investigador el poder identificar una causa y efecto entre las variables.

Las variables categóricas y continuas

Las variables categóricas se conocen también como variables cualitativas o factores. Las variables categóricas pueden clasificarse ya sea como nominales, dicotómicas u ordinales.

- Las variables nominales son variables que tienen dos o más categorías, pero que no tienen un orden intrínseco. Es de destacar que las diferentes categorías de una variable nominal también puede ser referidas como grupos o niveles de la variable nominal. Ejemplo: clasificar los vehículos según su color.
- Las variables dicotómicas son variables nominales que tienen sólo dos categorías o niveles. Ejemplo: clasificar personas según su género.
- Las variables ordinales son variables que tienen dos o más categorías al igual que las variables nominales sólo que las categorías también pueden ser ordenadas o clasificadas (ya sean variables numéricas o alfabéticas). En este caso también es preciso destacar que las diferentes categorías de una variable ordinal también puede ser referidas como grupos o niveles. Ejemplo: grado de satisfacción con un producto.

Las variables continuas se conocen también como variables cuantitativas. Las variables continuas pueden clasificarse sea como variables de intervalo o de relación.

- Las variables de intervalo son variables en las que su característica principal es que se puede medir a lo largo de un continuo y tienen un valor numérico. Ejemplo: temperatura medida en grados Celsius.
- Las variables de relación son variables de intervalo, pero con la condición añadida de que el cero de la medición indica la no existencia de esa variable. Ejemplo: temperatura medida en Kelvin.

Análisis usual de los resultados en campañas experimentales

Para cada una de las variables que pretenden ser analizadas en el presente estudio (Resistencias mecánicas, Microdureza, Consistencia, Contenido de aire, Absorción, Densidad y Porosidad) se procede a la realización de un análisis estadístico de los datos en profundidad, intentando así huir de interpretaciones simplistas que puedan llevar a errores en la interpretación y discusión de los resultados.

Normalmente, cuando se trata de estudios de laboratorio se suele trabajar con entre tres y seis resultados de la variable de interés (según el ensayo realizado) y con su media aritmética. Con esto, si uno de los resultados de todas las determinaciones difiere de la media en ± 10%, se desecha este resultado y se calcula la media de los resultados restantes. Si un nuevo resultado dentro de estas determinaciones restantes difiere de su media en más de ± 10%, se desecha toda la serie; volviendo a realizarse la fabricación de las muestras y los ensayos pertinentes.

Como se ha indicado, para la interpretación de los resultados, en el presente trabajo se ha optado por un análisis más sofisticado y menos simplista, con el que se tomaran una serie de hipótesis y mediante el uso de la estadística se aprobarán o rechazaran esas hipótesis, que serán las que nos permitirán sacar conclusiones robustas (Lund & Lund, 2013). Las pruebas estadísticas elegidas han sido:

- prueba T para muestras independientes, en el caso en que tengamos un factor con únicamente dos niveles posibles.
- prueba ANOVA de un factor, cuando en un mismo factor existan tres o más niveles de variación.

Y se han realizado con la ayuda del programa estadístico SPSS con licencia proporcionada por la Universitat Politècnica de Catalunya.

Prueba T para muestras independientes

La prueba T para muestras independientes es una prueba estadística utilizada para determinar si la media de una variable dependiente es la misma cuando la medimos en dos grupos diferentes. Exactamente, con la prueba T para muestras independientes lo que se hace es determinar si la diferencia de las medias de ambos grupos es estadísticamente diferente de cero.

Para llevar a cabo una prueba T para muestras independientes es necesario:

- Una variable independiente categórica (o factor) con dos grupos (o niveles). Los grupos deben ser independientes, esto es, que ninguna de las muestras pueda estar en ambos niveles del factor a la vez.
- Una variable dependiente continua, que es la que queremos estudiar.

Lo que intentamos con esta prueba, es determinar si la media de la población es igual para los dos grupos independientes considerados. La palabra población es importante porque la muestra que hemos recogido supone reflejar la población en la que estamos interesados. El propósito de la prueba T para muestras independientes es el de determinar cuando la media de la población de los dos grupos es diferente, y no sólo una consecuencia de la variación natural del muestreo. Esto es debido a que bajo circunstancias normales, calcular la diferencia de medias de la variable dependiente entre grupos es simple; simplemente restando la media de la variable dependiente en un grupo y la media de la variable dependiente del otro grupo, la cual cosa nos dará la diferencia de medias entre grupos. Sin embargo, esto sólo nos da la diferencia de medias entre grupos para la muestra seleccionada. Esto significa que esta diferencia de medias entre grupos naturalmente variará para las diferentes muestras que podamos elegir. El valor de la prueba T para muestras independientes es el que se considera para la población, es decir, estaremos analizando si la media de la población de ambos grupos es diferente, y no solo la de la muestra elegida (Pérez López, 2009).

Prueba ANOVA de un factor

La prueba ANOVA de un factor para el análisis de la varianza es una extensión de la prueba T para muestras independientes y se utiliza para determinar si existe diferencia estadística entre la media de la población de tres o más niveles de una misma categoría.

Para llevar a cabo una prueba ANOVA de un factor es necesario disponer de:

- Una variable independiente categórica (o factor) con tres niveles o más. Los grupos deben ser independientes, esto es que ninguna de las muestras pueda estar en varios niveles del factor a la vez.
- Una variable dependiente continua.

Lo que intentamos con esta prueba, es determinar si la media de la población es igual para todos los grupos independientes considerados, o existen diferencias entre ellos, aunque sólo sea entre algunos. La palabra población es importante porque la muestra que hemos recogido supone reflejar la población en la que estamos interesados. El propósito de la prueba ANOVA de un factor es el de determinar cuando la media de la población de los todos los grupos es diferente, y no sólo una consecuencia de la variación natural de muestreo. Esto es debido a que bajo circunstancias normales, calcular la diferencia de medias de la variable dependiente entre grupos es simple; simplemente restando la media de la variable dependiente en un grupo y la media de la variable dependiente del otro grupos, la cual cosa nos dará la diferencia de medias entre grupos. Sin embargo, esto sólo nos da la diferencia de medias entre grupos para la muestra seleccionada. Esto significa que esta diferencia de medias entre grupos naturalmente variará para las diferentes muestras que podamos elegir. El valor de la prueba ANOVA de un factor es el que se considera para la población, es decir, estaremos analizando si la media de la población de todos los grupos es diferente, y no solo la de la muestra elegida.

Conseguir un resultado estadísticamente significativo para un test ANOVA de un factor indica que no todas las medias son iguales, pero no informa dónde están las diferencias. Es por ese motivo que, posteriormente, y en caso de que sí exista diferencia entre las medias de los diferentes grupos, se procederá a realizar un estudio para determinar entre qué grupos existe una diferencia estadísticamente significativa y en cuales no, contemplando todas las combinaciones posibles y posibles diferencias entre cada uno de los niveles. Se realizará un tratamiento a posteriori Tukey en caso de obtener igualdad de varianzas y un tratamiento a posteriori Games-Howell en caso de obtener desigualdad de varianzas (Visauta Vinacua & Martori i Cañas, 2005). Ejecutando estas pruebas a posteriori, se comparará estadísticamente todos los grupos posibles en referencia a la variable dependiente. Dependiendo de los resultados de la prueba de homogeneidad de varianzas, se ignorará la prueba que no sea necesaria .

Tamaño de la muestra y equilibrio/desequilibrio de diseños

Como norma general, un estudio debe tener seis o más muestras en cada grupo para poder realizar una prueba T para muestras independientes o una prueba ANOVA de un factor, aunque lo ideal sería tener más. Ambas pruebas pueden ser realizada con menos de seis participantes, pero la capacidad de inferir o de generalizar estos resultados a una población más grande será más difícil.

Si se dispone del mismo número de muestras en cada grupo, se dispone de lo que se denomina un diseño "equilibrado". Alternativamente, si el tamaño de la muestra no es el mismo para ambos grupos, se dispondrá de un diseño "desequilibrado". Generalmente, cuanto más desequilibrado es un diseño, mayor es el efecto negativo de violar una suposición en la validación de la prueba. Idealmente, se desea siempre disponer de un diseño equilibrado, y esto es lo que se ha tenido en cuenta a la hora de planificar los ensayos necesarios para el presente estudio.

Verificaciones a realizar antes de la realización de las pruebas

En un estudio de investigación, el error tipo I también denominado error de tipo alpha (α) o falso positivo, es el error que se comete cuando el investigador no acepta la hipótesis nula (H_0) siendo esta verdadera en la población. Es equivalente a encontrar un resultado falso positivo, porque el investigador llega a la conclusión de que existe una diferencia entre las hipótesis cuando en realidad no existe. Se relaciona con el nivel de significancia estadística.

La hipótesis de la que se parte (la hipótesis nula H_0) aquí es el supuesto de que la situación experimental presentaría un "estado normal". Si no se advierte este "estado normal", aunque en realidad existe, se trata de un error estadístico tipo I.

Además, en un estudio de investigación, también puede existir el error de tipo II, también llamado error de tipo beta (β) o falso negativo, donde β es la probabilidad de que exista este error. Este tipo de error se comete cuando el investigador no rechaza la hipótesis nula siendo esta falsa en la población. Es equivalente a la probabilidad de un resultado falso negativo, ya que el investigador llega a la conclusión de que ha sido incapaz de encontrar una diferencia que existe en la realidad.

En el presente estudio, por defecto, se van a utilizar siempre intervalos de confianza del 95%, es decir, se determinarán un par de valores entre los cuales se estima que estará cierto valor desconocido con una probabilidad del 95% de acierto. Esto equivale a declarar la significancia estadística al nivel de $p < 0,05$. Obviamente, si se desea, esto puede cambiarse e introducir cualquier valor entre 1 y 99. Por ejemplo, un intervalo de confianza del 99% equivaldría a declarar significancia estadística al nivel de $p < 0,01$. A fin de no aumentar el error tipo II y de mantener un valor relativamente coherente para el error tipo I, se fija la confianza al 95% (valor comúnmente utilizado en estudios estadísticos).

A fin de obtener unos resultados válidos mediante la realización tanto de una prueba T para muestras independientes como de una prueba ANOVA de un factor, es necesario verificar que se cumplen las siguientes cuatro suposiciones antes de ejecutar la prueba.

1- Independencia de las observaciones

2- No existen valores atípicos en ninguno de los grupos

3- Las datos de la muestra siguen una distribución normal

4- Existe homogeneidad de varianzas

Figura 1. Comprobaciones a realizar sobre los datos antes de iniciar un análisis estadístico.

1- <u>Independencia de las observaciones.</u>

Esta suposición radica en que no hay relación alguna entre las muestras de cada grupo o entre las muestras de los diferentes grupos. Esto significa que debe haber diferentes muestra en cada grupo, sin muestras que pertenezcan a diferentes grupos a la vez. La independencia de las observaciones o de las muestras es en gran medida una cuestión de diseño del estudio.

2- <u>No existen valores atípicos en ninguno de los grupos.</u>

Para los diferentes grupos, si hay resultados que resultan inusuales para un grupo concreto, caso de que su valor sea extremadamente pequeño o extremadamente grande comparado con los demás resultados, estos valores se denominan puntos atípicos. Los valores atípicos pueden tener un gran efecto negativo en los resultados, ya que pueden ejercer una influencia grande en la media y en la desviación estándar para ese grupo, que puede afectar a los resultados de las pruebas estadísticas. Los valores atípicos son más importantes a considerar cuando se tienen pequeños tamaños de muestra (nuestro caso), ya que el efecto de los valores atípicos será mayor. Debido al efecto que los valores atípicos pueden tener en los resultados, es necesario elegir si se desea incluir estos datos en las muestras cuando se realizan las pruebas estadísticas.

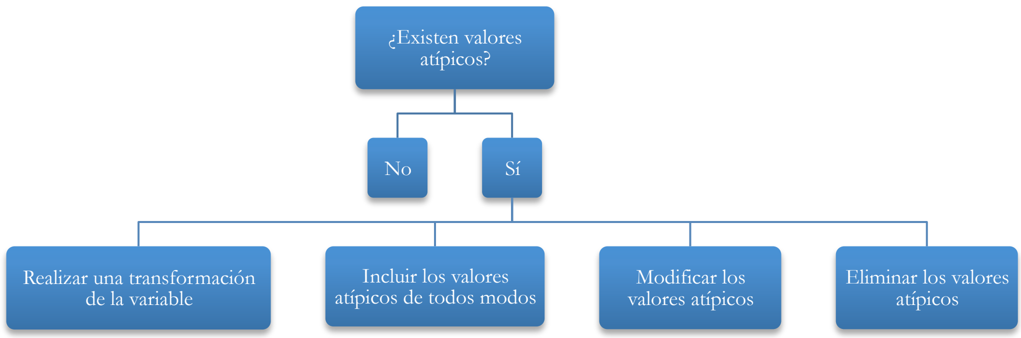

Figura 2. Pasos a seguir en la comprobación de la existencia y tratamiento de valores atípicos en los datos.

Existen varios métodos que pueden ser utilizados para detectar la existencia de valores atípicos en cada grupo, pero en el presente estudio se usarán los diagramas de caja para identificar la existencia o no existencia de valores atípicos. Se opta por este método ya que el hecho de usar diagramas de caja es una metodología fácil, rápida y sencilla para detectar valores atípicos

Al observar el diagrama de cajas generado al analizar la variable dependiente frente al factor de diferentes posibles niveles, se puede comprobar que todos los valores que se encuentren a una distancia mayor de 1,5 veces la longitud de la caja desde el borde de su caja, se clasifican como valores atípicos y se ilustran como puntos circulares (o). Todos los valores que se encuentren a una longitud mayor de 3 veces la longitud de caja de distancia desde el borde de su caja se clasifican como puntos extremos, y se ilustran con un asterisco (*). Estos dos tipos de valores atípicos se encuentran etiquetados con su número de caso, para su fácil identificación en el listado de datos. De modo que, si no hay valores atípicos en los datos, esto se evidencia por la ausencia de puntos circulares y/o asteriscos. En los gráficos representados a continuación puede observarse que, en el primero de los casos, no existen valores atípicos en los datos, como evidencia la ausencia de puntos circulares o asteriscos. En el segundo de los casos se muestra como es un diagrama de cajas con puntos atípicos y extremadamente atípicos en la serie de datos analizados.

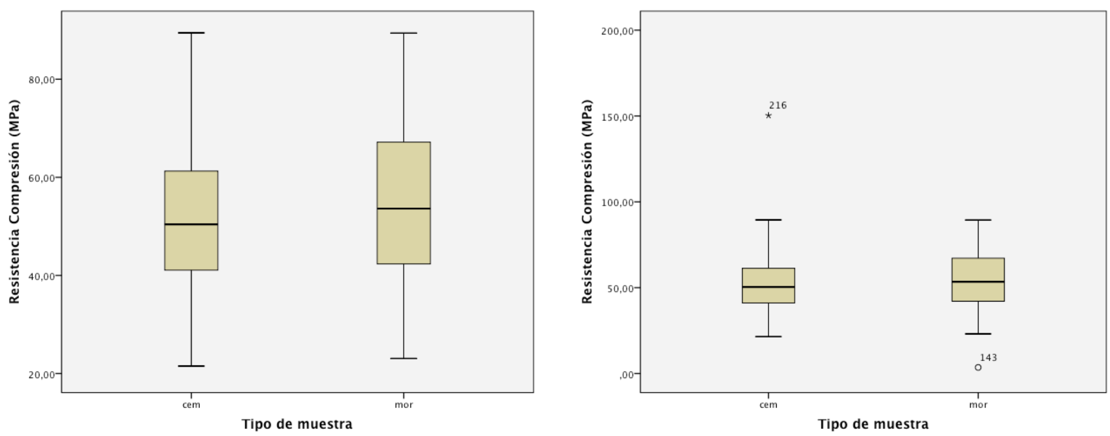

Figura 3. Ejemplos de conjunto de datos sin valores atípicos (a) y conjunto de datos con un valor atípico y un valor extremo (b).

Basándonos en los resultados recién obtenidos, pueden suceder dos cosas completamente opuestas: se han detectado puntos atípicos o no se han detectado puntos atípicos en los resultados de nuestras muestras. Si no se observan puntos atípicos en el diagrama de caja podemos pasar directamente a comprobar si se cumple la siguiente condición, esta es si los datos de las muestras se encuentras normalmente distribuidos. Si no es así, y se han detectado valores atípicos es necesario saber de que tipo de valores atípicos se trata y como tratarlos. En general, existen tres motivos por los que podemos encontrar valores atípicos en nuestros datos: errores en la entrada de datos, errores de medida, y valores realmente inusuales.

- *Errores en la entrada de datos:* la primera consideración a tener en cuenta debe ser la de comprobar si los valores extremos son el resultado de errores en la entrada de datos. Un error en el ingreso de los datos se produce cuando simplemente se ha tecleado un valor erróneo a la hora de pasar los datos al programa de análisis estadístico. Este es el tipo de valor atípico más fácil de solucionar. Si alguno de los valores atípicos se debe a un error en la entrada de los datos, entonces simplemente debe reemplazarse con el valor correcto y volver a ejecutar todas las pruebas de hipótesis. Una vez rectificado ese valor, es necesario volver a ejecutar todas estas pruebas ya que a pesar de la modificación de ese dato podría ser que se siguiera tratando de un valor atípico, o aún peor, podría llevar a otros datos de la muestra a ser clasificados ahora como valores atípicos.

- *Errores de medida:* si nos encontramos con que los valores atípicos no se deben a errores en la entrada de datos, lo próximo debe ser considerar que se tratan de errores de medida (por ejemplo, debido a un mal funcionamiento del equipo de medida o que se trate de valores fuera del rango donde deberían encontrarse todos los datos). Los errores de medición a menudo no son corregibles y normalmente estos deberán ser eliminados del análisis. A veces, sin embargo, si se ha generado un valor fuera de rango y se conoce en que dirección se ha salido del rango, puede reemplazarse ese valor con el mayor valor válido. Esto puede ser mejor que no tener ningún valor introducido al respecto, pero puede conducir a sesgos. Debe recordarse, que si se hace la corrección ni que sea de un solo dato, se debe de volver a ejecutar todas las pruebas de hipótesis, por los posibles cambios que esta modificación pueda ocasionar sobre los resultados de las pruebas.

- *Valores realmente inusuales:* si se ha establecido que un valor atípico no es el resultado ni de un error de entrada de datos ni de un error de medición, lo más probable es que se trate de un valor realmente inusual. Estos son los datos más difíciles de tratar porque a pesar de que no son los ideales desde el punto de vista estadístico (es decir, violan uno de los supuestos de las pruebas estadísticas), no hay ninguna buena razón para rechazar como no válido. Lamentablemente, lo que se puede hacer en esta situación es controvertido y no existe un único procedimiento recomendado, pero esto no significa que no existan opciones de tratarlo. Como nota importante, si se disponen de múltiples valores atípicos y se está considerando la posibilidad de eliminarlos del análisis, quítese el valor atípico más extremo y luego vuélvase a ejecutar la prueba. A veces, los demás valores atípicos dejan de serlo cuando el valor atípico más extremo es eliminado del análisis.

De modo que una vez llegados a este punto podemos tomar cuatro soluciones diferentes:

- Mantener los valores atípicos: en caso de que no se desee eliminar un valor atípico o se crea que este no puede ser eliminado del análisis, se contemplan tres opciones sobre la manera de proceder:

- *Modificar el valor atípico sustituyéndolo por uno que sea menos extremo*, por ejemplo, el siguiente valor más grande en su lugar. La razón por la que se hace el valor ligeramente mayor que el segundo valor más grande es la de mantener su rango, es decir, este valor atípico modificando sigue siendo el valor más grande del grupo. La principal ventaja de esta solución es que se mantiene parte de la información contenida en el valor atípico (se tiene un dato y además es el valor más grande), pero la mayor desventaja es que los datos ahora parecen ser más normales (es decir, menos dispersos) de lo que realmente son, y la modificación de estos valores genuinos podría introducir un sesgo en el análisis.

- *Realizar un cambio de variable sobre la variable dependiente*. La transformación de los datos puede ser una opción válida, ya que puede dar lugar a que los valores atípicos que inicialmente se determinen como desproporcionados debido al reducido tamaño de la muestra, después de esta transformación, ya no se clasifiquen como valores atípicos. Sin embargo, las transformaciones no suelen garantizar esta exclusión de los valores atípicos, a menos que los datos se encuentren inicialmente distribuidos de forma no-normal, por lo que sólo debería considerarse esta opción si también es necesario aplicar una transformación debido a la no-normalidad. Este método también viene con la desventaja de hacer que la interpretación de las pruebas estadísticas sean más difíciles de interpretar en la medida en que ya no se están utilizando los valores de los datos originales, debiendo de deshacer el cambio de variables realizado cada vez que se quieran interpretar los resultados. Si se opta por esta opción, se deberá volver a ejecutar todas las pruebas de hipótesis con la nueva variable transformada.

- *Incluir el valor atípico en el análisis de todos modos*, si se cree que el resultado obtenido no resultará significativamente afectado por este valor atípico (procedimiento que debería ser acompañado de un análisis de sensibilidad). Esta opción requiere mucha más confianza por parte de la persona que está realizando el análisis estadístico de los datos, pero puede ser una estrategia perfectamente aceptable en el tratamiento de los valores atípicos. Idealmente, lo que se está pretendiendo es encontrar un método que evalúe si el valor atípico tiene un efecto apreciable en el análisis. Un método que puede utilizarse es el de realizar una prueba estadística con y sin el valor atípico incluido en el análisis. A continuación, comparar los resultados y decidir si los dos resultados difieren lo suficiente como para llegar a conclusiones diferentes. Si las conclusiones son esencialmente las mismas (por ejemplo, ambos análisis dan lugar a un resultado estadísticamente significativo), es posible mantener el valor atípico en los datos.

- *Eliminar los valores atípicos*: si se prefiere, pueden quitarse los valores típicos del análisis. Si se elige este método, se debe proporcionar información sobre los datos que se han retirado del análisis. Esto es para que el lector pueda formarse un juicio preciso sobre por qué se han quitado y cómo podrían haber afectado estos al análisis.

3- <u>Las datos de la muestra siguen una distribución normal.</u>

La suposición de normalidad es necesaria para las pruebas de significancia obtenidas al realizar una prueba estadística. Hay muchos métodos diferentes disponibles para probar esta hipótesis de normalidad, y normalmente se distingue entre métodos gráficos y métodos numéricos.

Sin embargo, cabe destacar que las prueba estadísticas consideradas, estas son la prueba T para muestras independientes y la prueba ANOVA de un factor, se consideran pruebas "robustas" frente a violaciones de la normalidad. Esto significa que violaciones de esta suposición pueden ser toleradas y las prueba seguirán proporcionando resultados válidos. Por lo tanto, a menudo se oye hablar de que estas pruebas solamente requieren unos datos aproximadamente normales. Además, a medida que aumenta el tamaño de la muestra, la distribución puede ser muy no-normal, y gracias al Teorema del Límite Central, estas pruebas estadísticas puede proporcionar resultados válidos. Además cabe señalar que si las distribuciones son sesgadas de una manera similar (por ejemplo, ambas con moderada asimetría negativa), esto no es tan problemático como cuando se comparan grupos que tienen diferentes distribuciones (por ejemplo, la distribución del grupo A tiene una moderada asimetría positiva, mientras que la distribución del grupo B tiene una moderada asimetría negativa).

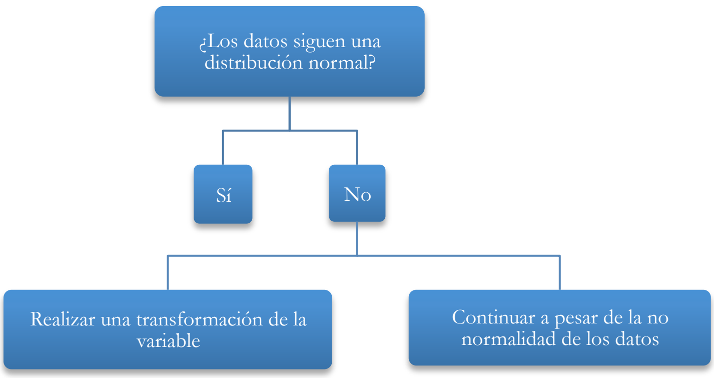

Figura 4. Pasos a seguir en la comprobación de la existencia de normalidad en los datos.

Existen varios métodos que pueden ser utilizados para determinar si las muestras que forman cada uno de los grupos se encuentran normalmente distribuidas. En el presente estudio se emplearán dos métodos diferentes: uno analítico y otro gráfico. Con esta complementariedad de métodos, seremos capaces de determinar con mayor exactitud si los participantes de cada grupo se encuentran distribuidos de forma normal o no. De modo que como método numérico se utilizará la prueba de Shapiro-Wilk para normalidad, y como método gráfico se utilizará el gráfico Q-Q normal.

Prueba de Shapiro-Wilk para la normalidad

La prueba de Shapiro-Wilk se recomienda cuando se tiene tamaños de muestra pequeños (<50 participantes) y no se está muy seguro de la interpretación visual de la gráfica Q-Q normal u otros métodos gráficos utilizados para la prueba de normalidad. La prueba de Saphiro-Wilk prueba si los datos se distribuyen normalmente para cada grupo de la variable independiente. Por lo tanto, habrá tantas pruebas de Shapiro-Wilk como niveles de la variable independiente.

Con el fin de comprender si los resultados de cada grupo se distribuyen normalmente, es necesario consultar el nivel de significancia (abreviado como "Sig") de la prueba para uno de los grupos.

La hipótesis nula de la prueba de Shapiro-Wilk es que la distribución de sus datos es igual a una distribución normal y la hipótesis alternativa es que la distribución de sus datos no es igual a una distribución normal.

H_0: *la distribución de los datos es igual a una distribución normal*

H_A: *la distribución de los datos no es igual a una distribución normal*

Por lo tanto si se rechaza la hipótesis nula ($p < 0,05$), esto significa que la distribución de los datos no es igual a una distribución normal y si no se puede rechazar la hipótesis nula, los datos se distribuyen normalmente. Simplificadamente, también puede utilizarse la siguiente interpretación: si los datos se distribuyen normalmente (es decir, el supuesto de normalidad se cumple), el nivel de significancia debe ser superior a 0,05 (es decir, $p > 0,05$). Si los datos no se distribuyen normalmente, es decir, el supuesto de normalidad es violado, el nivel de significancia será inferior a 0,05 (es decir, $p < 0,05$).

Pruebas de normalidad

	Tipo de muestra	Shapiro-Wilk		
		Estadístico	gl	Sig.
Resistencia Compresión (MPa)	cem	,979	156	,116
	mor	,981	156	,203

Figura 5. Ejemplo de resultado obtenido al realizar la prueba de Shapiro-Wilk para la normalidad.

En la figura anterior, se puede comprobar que ambos niveles de significancia son mayores que 0,05 (son 0,116 para el grupo cemento y 0,203 para el grupo mortero), y por lo tanto, la variable dependiente "Resistencia a Compresión (MPa)", tiene una distribución normal para cada nivel de la variable independiente "Tipo de muestra".

Gráfico Q-Q normal

Si los tamaños de las muestras son mayores de 50, probablemente resulte beneficioso el uso de métodos gráficos, como un gráfico Q-Q normal porque en tamaños de muestra más grandes, la prueba de Shapiro-Wilk señala incluso las desviaciones de menor importancia como estadísticamente significativas (es decir, obteniendo que los datos no están normalmente distribuidos cuando realmente si lo están). Por este motivo, se opta por utilizar ambos métodos simultáneamente ya que esta complementariedad de métodos nos permite estar seguros de los resultados obtenidos para muestras grandes y pequeñas.

El gráfico Q-Q normal es uno de los mejores métodos visuales de evaluación de la normalidad. Si los datos se distribuyen normalmente, los puntos circulares que representan los datos se colocan aproximadamente a lo largo de la línea diagonal del gráfico Q-Q normal. Cabe destacar la palabra aproximadamente en la afirmación anterior, ya que, en realidad, siempre habrá alguna variación entre los datos y la diagonal, incluso cuando los datos se encuentren distribuidos de forma normal.

Se puede observar en los gráficos representados a continuación, que los puntos están aproximadamente distribuidos a lo largo de la línea diagonal, tanto para muestras de tipo cemento como para muestra de tipo mortero; por lo que se puede concluir que los datos analizados en ambos grupos se distribuyen normalmente (o al menos una distribución aproximadamente normal, lo cual resulta suficiente para la mayoría de pruebas paramétricas).

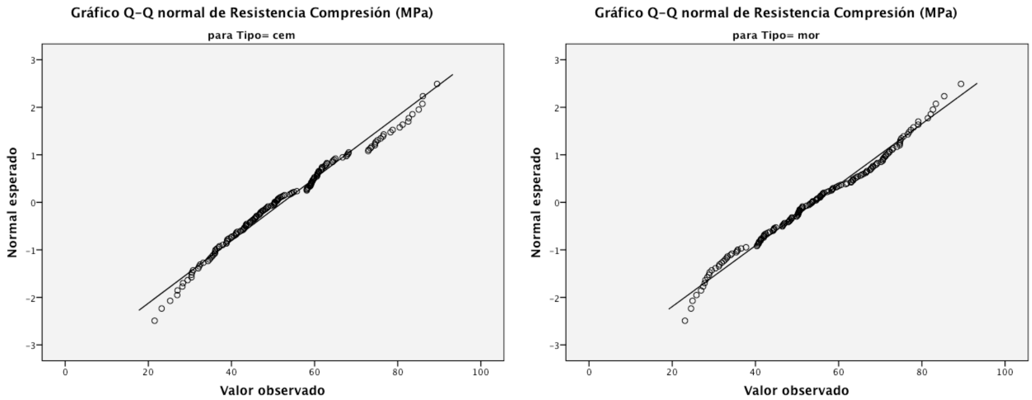

Figura 6. Ejemplo de gráficos Q-Q normales obtenidos al ejecutar la prueba de normalidad.

Estos resultados concuerdan con los resultados obtenidos mediante la prueba analítica de Saphiro-Wilk para la normalidad, que ya nos indicaban que los datos seguían una distribución normal para los dos niveles del mismo grupo. De modo que, como se ha comentado anteriormente, se seguirá utilizando ambos métodos para que, en las ocasiones en las que pueda surgir algún tipo de duda, esta complementariedad de métodos pueda ayudarnos a aclarar las dudas con respecto a la normalidad.

En función de los resultados obtenidos de las pruebas de normalidad, existen dos opciones diferentes a la hora de proceder:

- Si los datos siguen una distribución normal, puede pasarse a la siguiente verificación sobre los datos (homogeneidad de varianzas).

- Si los datos no se distribuyen normalmente, entonces se dispone de dos opciones a la hora de proceder:

    - *Realizar una transformación de variable:* puede realizarse una transformación de la variable dependiente que está siendo objeto de estudio para que, con suerte, estos datos transformados se distribuyan normalmente. Aunque esto puede resultar ser un método fácil con el cual conseguir datos distribuidos normalmente, no siempre funciona porque a veces no se consiguen hallar esa transformación que funcione. Además, los resultados transformados pueden resultar más difíciles de interpretar, ya que estos no representan ni tienen mucho que ver con los valores originales.

    - *Continuar independientemente de haber obtenido violación de la normalidad:* esta opción contempla el seguir adelante y ejecutar la prueba estadística pertinente a pesar de no tener datos distribuidos normalmente ya que cualquiera de las dos pruebas elegidas resultan ser pruebas bastante robustas a las desviaciones de la normalidad. En general, si el tamaño de la muestra en cada grupo es igual, o prácticamente igual, sólo violaciones fuertes de normalidad podrían causar problemas. Es más, si los tamaños de muestra no son pequeños, incluso para distribuciones bastante sesgadas (siempre y cuando los grupos tengas aproximadamente la misma asimetría) no se suelen presentar problemas en este aspecto. En conclusión, la no-normalidad no afecta sustancialmente a la tasa de error tipo I y tanto la prueba T para muestras independientes como la prueba ANOVA de un factor pueden ser consideradas unas pruebas robustas. Sin embargo, si se elige este camino, se debe reportar esta violación en los resultados.

4- Existe homogeneidad de varianzas

La suposición de homogeneidad de varianzas expresa que la varianza de la población de cada grupo es la misma. Esto se puede comprobar gráficamente o mediante la ejecución de la prueba de Levene para la igualdad de varianzas. Si el tamaño de la muestra en cada grupo es similar, la violación de este supuesto no es a menudo demasiado grave. Sin embargo, si los tamaños de las muestras son muy diferentes, la prueba T para muestras independientes es sensible a la violación de este supuesto.

El hecho de que esta hipótesis de homogeneidad de varianzas se cumpla o sea violada, afecta a como se deben interpretar los resultados obtenidos mediante la ejecución de la prueba T para muestras independientes. Esta suposición afirma que los datos de la variable dependiente en ambos grupos tienen varianzas iguales en la población. Si esto no es así, es decir, las varianzas son desiguales, esto puede afectar a la tasa del error tipo I. Por lo tanto, se desea que la varianza de la población de la variable dependiente sea igual para ambos niveles de la variable independiente.

Para comprobar si hay igualdad en las varianzas, es necesario consultar el nivel de significancia de la prueba de Levene para la igualdad de varianzas, como se muestra en la siguiente figura.

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas	
		F	Sig.
Resistencia Compresión (MPa)	Se han asumido varianzas iguales	,177	,675
	No se han asumido varianzas iguales		

Figura 7. Ejemplo de resultado obtenido al realizar la prueba de Levene para la igualdad de varianzas.

La hipótesis nula de la prueba de Levene para la igualdad de varianzas es que la varianza en la población de ambos grupos es igual (es decir, existe homogeneidad de varianzas o homocedasticidad) y la hipótesis alternativa es que la varianza en la población de ambos grupos es diferente.

H_0: *las varianzas poblacionales son iguales en ambos grupos*

H_A: *las varianzas poblacionales son diferentes en ambos grupos*

Simplificadamente, puede decirse que si las varianzas de las poblaciones de ambos grupos son iguales, esta prueba devolverá un valor de significancia superior a 0,05 (es decir, $p > 0{,}05$), lo que indica que se ha cumplido con el supuesto de homogeneidad de varianzas. Sin embargo, si la prueba devuelve un valor de significancia inferior a 0,05 (es decir, $p < 0{,}05$), las varianzas no son iguales y consecuentemente se habrá violado el supuesto de homogeneidad de varianzas. En este ejemplo, el valor de la significancia es 0,675 (es decir, $p = 0{,}675$), la cual cosa significa que las varianzas de las poblaciones de los datos para los dos grupos son iguales. Por lo tanto, la hipótesis de homogeneidad de varianzas se cumple.

Realización de la prueba T para muestras independientes

Una vez se han realizado las cuatro comprobaciones sobre los datos, puede realizarse la prueba T para muestras independientes, estando ahora seguros de que los resultados que obtengamos tendrán sentido.

El primer paso tras la ejecución de la prueba T para muestras independientes, es la de interpretar los estadísticos de grupo, que contienen información útil para interpretar los resultados de ambos grupos. Los estadísticos de grupo muestran información sobre el número de casos analizados, la media, la desviación estándar y el error estándar sobre la media, y se recogen, de forma sintetizada, en una tabla como la siguiente:

Estadísticos de grupo

	Tipo de muestra	N	Media	Desviación típ.	Error típ. de la media
Resistencia Compresión (MPa)	cem	156	52,3083	15,22237	1,21877
	mor	156	54,2168	15,59909	1,24893

Figura 8. Ejemplo de tabla de estadísticos de grupo obtenida al realizar la prueba T para muestras independientes.

Cada fila de la tabla anterior presenta estadísticos sobre la variable dependiente, para las diferentes categorías de la variable independiente. Esta tabla puede ser usada para entender ciertos aspectos de los datos, como por ejemplo, si hay un número igual de participantes en cada uno de los grupos, que grupo ha obtenido la mayor o menor puntuación para la variable dependiente (y lo que eso significa para el estudio en cuestión), y si la variación en cada grupo es similar.

Ahora que ya se pose de una idea general de los datos (gracias a las interpretaciones realizadas mediante el uso de los estadísticos de grupo), es necesario determinar el tamaño o magnitud de la diferencia entre los dos grupos y si esta diferencia de medias es estadísticamente significativa o no. Esta determinación dependerá del resultado obtenido en la prueba de Levene para la igualdad de varianzas.

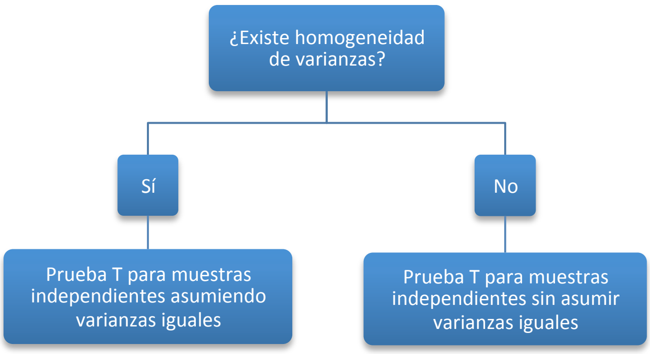

Figura 9. Esquema sobre como actuar en caso de tener o no tener homogeneidad de varianzas al realizar una prueba T para muestras independientes.

Con lo que llegados a este punto podemos seguir 2 caminos diferentes, según si se ha asumido varianzas iguales o no:

- La hipótesis de igualdad de varianzas es cierta. Esta suposición afirma que los valores de la variable dependiente en ambos grupos tienen varianzas iguales en la población. Si realmente no fuera así, y las varianzas de la población realmente no fueran iguales, se estaría cometiendo un error tipo I.
  Cuando esta hipótesis se cumple, es necesario interpretar la fila llamada "Se han asumido varianzas iguales" dentro de la prueba T para muestras independientes con el fin de obtener los resultados apropiados.

Prueba de muestras independientes

		Prueba T para la igualdad de medias						
							95% Intervalo de confianza para la diferencia	
		t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Resistencia Compresión (MPa)	Se han asumido varianzas iguales	-1,094	310	,275	-1,90853	1,74505	-5,34217	1,52512
	No se han asumido varianzas iguales	-1,094	309,815	,275	-1,90853	1,74505	-5,34218	1,52513

Figura 10. Ejemplo de cómo interpretar la tabla de resultados de la prueba T para muestras independientes en caso de asumir varianzas iguales.

Esta fila utiliza las ecuaciones estándar (con varianza combinada) para calcular los resultados de la prueba T para muestras independientes.

Al determinar si la diferencia de medias entre los dos grupos es estadísticamente significativa, se considera de utilidad y así se hará, mostrar el intervalo de confianza al 95% de esa diferencia de medias. Este resultado estará informando sobre la diferencia de medias y estando seguros al 95%, entre que intervalo se encuentra esta.

Ahora que ya se ha proporcionado información sobre la magnitud o tamaño de la diferencia, es necesario determinar si precisamente la media de estos dos grupos resulta estadísticamente diferente. Esta información es el resultado que nos proporciona la realización de la prueba T para muestras independientes, pero para entender lo que se nos está mostrando, es necesario comprender que hipótesis nula y que hipótesis alternativa se está contrastando en la presente prueba. Una prueba T para muestras independientes toma como hipótesis nula que la media de las 2 poblaciones es igual y como hipótesis alternativa que la media de las 2 poblaciones es diferente:

H_0: la media de las poblaciones de los 2 grupos son iguales ($\mu_1 = \mu_2$)

H_A: la media de las poblaciones de los 2 grupos no son iguales ($\mu_1 \neq \mu_2$)

También se puede reinterpretar las hipótesis recién citadas para reflejar las diferencias entre medias. De modo que la hipótesis nula ahora es:

H_0: la diferencia entre las medias de las poblaciones de los 2 grupos es cero ($\mu 1 - \mu 2 = 0$)

Y la hipótesis alternativa reinterpretada es:

H_A: la diferencia entre las medias de las poblaciones de los 2 grupos no es cero ($\mu 1 - \mu 2 \neq 0$)

La prueba T para muestras independientes calcula un nivel de significancia (p-valor), que es la probabilidad de que las medias de las muestras de los grupos analizados sean al menos tan diferentes como las analizadas en el estudio, dando por válida la hipótesis nula. Si la probabilidad es suficientemente pequeña (asumiremos $p < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con lo que se puede concluir que las medias de los 2 grupos son diferentes. De forma alternativa, se rechaza la hipótesis alternativa y por lo tanto no se puede rechazar la hipótesis nula si la probabilidad es alta (asumiremos $p > 0,05$), concluyendo que las medias de los 2 grupos son iguales. Cabe remarcar que no se acepta la hipótesis nula, sino que se afirma que ésta no puede ser rechazada. Es importante en este punto recordar que las hipótesis de la prueba hacen referencia a la media de las poblaciones, y no a las medias de las muestras analizadas. En el ejemplo que se muestra, $p = 0,275$ (es decir, $p > 0,05$). Por lo tanto, se puede concluir que la variable independiente "Resistencia Compresión (MPa)" es estadísticamente igual para ambos grupos de población, tanto para cementos como para morteros. O expresado de otra manera, la diferencia de medias en la resistencia a compresión entre las muestras de cemento y las de mortero no es estadísticamente significativa. Lo que esto significa es que hay una probabilidad del 27% de obtener una diferencia de medias al menos tan grande como el obtenido si la hipótesis nula es verdadera (la hipótesis nula que indica que no hay ninguna diferencia entre la media de ambos grupos). En este punto cabe recordar que la prueba T para muestras independientes determina si las medias de las poblaciones son iguales, y no las medias de las muestras elegidas y analizadas.

Existe un enlace entre el intervalo de confianza del 95% de la diferencia de medias y la significancia estadística de la diferencia de medias. Si el intervalo de confianza no contiene el número 0, esto significa que existe una diferencia significativa entre la diferencia de medias ($p < 0,05$). Si, por el contrario, el intervalo de confianza contiene el 0, no existe significancia estadística entre la diferencia de medias ($p > 0,05$).

Es importante recordar que el nivel de significancia (p-valor) no indica la fuerza o la importancia de la diferencia de medias entre los grupos, solamente indica la posibilidad de poder observar una diferencia de medias tan grande o mayor a la observada, dado que la hipótesis nula es cierta. Por ejemplo, si en el ejemplo anterior hubiéramos obtenido un p-valor de 0,137 ($p = 0,0137$), esto no significaría que fuera dos veces más fuerte o más importante que al haber obtenido $p = 0,026$. En términos simples, el p-valor simplemente trata de informar de si la diferencia entre los dos grupos estudiados no es un "golpe de suerte" y que realmente se esperaría ver una diferencia de medias como la del estudio en la población (y no sólo en la muestra). En este sentido, un p-valor menor indica simplemente el grado de confianza que se puede tener de que el resultado es un resultado verdadero. Si esa diferencia es importante o grande, eso no lo especifica el p-valor.

- La hipótesis de igualdad de varianzas es falsa. Cuando los resultados muestran que la hipótesis de homogeneidad de varianzas ha sido violada, es necesario consultar la fila inferior de la prueba T para muestras independientes.

Prueba de muestras independientes

		Prueba T para la igualdad de medias						
		t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia: Inferior	95% Intervalo de confianza para la diferencia: Superior
Resistencia Compresión (MPa)	Se han asumido varianzas iguales	-1,094	310	,275	-1,90853	1,74505	-5,34217	1,52512
	No se han asumido varianzas iguales	-1,094	309,815	,275	-1,90853	1,74505	-5,34218	1,52513

Figura 11. Ejemplo de cómo interpretar la tabla de resultados de la prueba T para muestras independientes en caso de no asumir varianzas iguales

Los resultados obtenidos en esta fila inferior utilizan la corrección de los grados de libertad de Welch-Satterthwaite y la corrección del cálculo con varianzas no agrupadas del estadístico T, los cuales tratan de corregir las varianzas desiguales.

A la hora de interpretar los resultados y de determinar si existe una diferencia significativa entre ambos grupos se procede de la misma manera que cuando se asumen varianzas iguales (las hipótesis que se contrastan en la prueba son las mismas), la única diferencia es que cuando no se han asumido varianzas iguales, se debe interpretar la fila de abajo, ya que se realiza una corrección de los datos teniendo en cuenta esta desigualdad en las varianzas.

Por último, como los grupos de la variable independiente no se encuentran en una escala continua, se procederá a la representación de los resultados mediante un gráfico de barras. En el gráfico de barras, se introducen las barras de error, que especifican los intervalos de confianza al 95%.

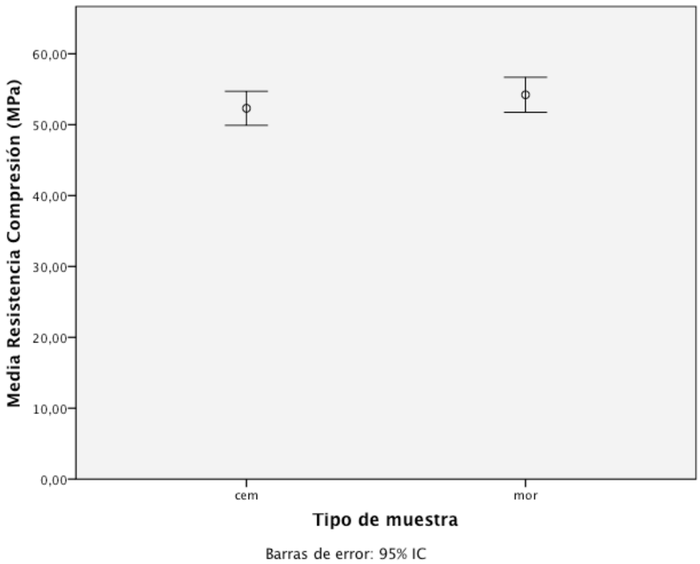

Figura 12. Ejemplo de gráfico de barras con intervalos de confianza del 95%.

Una vez se han realizado las cuatro comprobaciones sobre los datos, puede realizarse la prueba ANOVA de un factor, estando ahora seguros de que los resultados que obtengamos tendrán sentido.

Realización de la prueba ANOVA de un factor

El primer paso tras la ejecución de la prueba ANOVA de un factor, es la de interpretar los descriptivos, que contienen información útil para interpretar los resultados de ambos grupos. Los descriptivos muestran información sobre el número de casos analizados, la media, la desviación estándar, el error estándar sobre la media, un intervalo de confianza para la media del 95%, el valor mínimo y el máximo, y se recogen, de forma sintetizada, en una tabla como la siguiente:

Descriptivos

Resistencia Compresión (MPa)

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
3d	104	46,3137	13,17922	1,29233	43,7507	48,8768	21,51	71,81
7d	104	52,8770	14,26417	1,39872	50,1030	55,6510	25,82	83,54
28d	104	60,5968	15,40344	1,51043	57,6012	63,5924	31,89	89,46
Total	312	53,2625	15,41674	,87280	51,5452	54,9799	21,51	89,46

Figura 13. Ejemplo de tabla de descriptivos obtenida al realizar la prueba ANOVA de un factor.

Cada fila de la tabla anterior presenta estadísticos sobre la variable dependiente, para las diferentes categorías de la variable independiente. Esta tabla puede ser usada para entender ciertos aspectos de los datos, como por ejemplo, si hay un número igual de participantes en cada uno de los grupos, que grupo ha obtenido la mayor o menor puntuación para la variable dependiente (y lo que eso significa para el estudio en cuestión), y si la variación en cada grupo es similar.

Ahora que ya se pose de una idea general de los datos (gracias a las interpretaciones realizadas mediante el uso de los descriptivos), es necesario determinar el tamaño o magnitud de la diferencia entre los dos grupos y si esta diferencia de medias es estadísticamente significativa o no. Esta determinación dependerá del resultado obtenido en la prueba de Levene para la igualdad de varianzas.

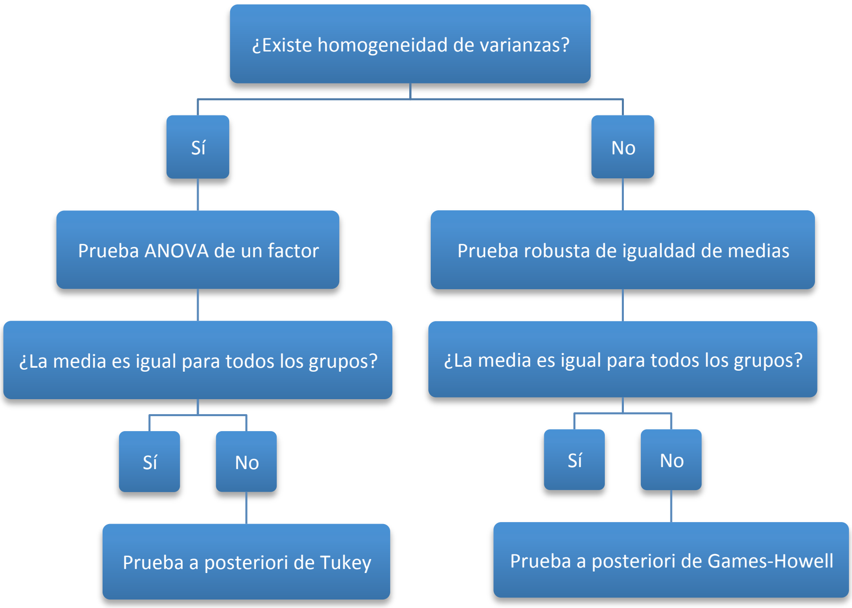

Figura 14. Esquema sobre como actuar en caso de tener o no tener homogeneidad de varianzas al realizar una prueba ANOVA de un factor.

Con lo que llegados a este punto podemos seguir dos caminos diferentes, según si se ha asumido varianzas iguales o no:

- La hipótesis de igualdad de varianzas es cierta. Esta suposición afirma que los valores de la variable dependiente en todos los grupos tienen varianzas iguales en la población. Si realmente no fuera así, y las varianzas de la población realmente no fueran iguales, se estaría cometiendo un error tipo I.
  Cuando esta hipótesis sobre a igualdad de varianzas se cumple, es necesario interpretar el nivel de significancia que nos devuelve la prueba ANOVA de un factor con el fin de obtener los resultados apropiados.

ANOVA de un factor

Resistencia Compresión (MPa)

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	10631,512	2	5315,756	25,955	,000
Intra-grupos	63285,686	309	204,808		
Total	73917,198	311			

Figura 15. Ejemplo de resultado obtenido al ejecutar la prueba ANOVA de un factor asumiendo igualdad de varianzas.

Ahora que ya se ha proporcionado información sobre la variable dependiente en todos sus niveles mediante los descriptivos, es necesario determinar si la media de todos estos grupos resulta estadísticamente diferente. Esta información es el resultado que nos proporciona la realización de la prueba ANOVA de un factor, pero para entender lo que se nos está mostrando, es necesario comprender que hipótesis nula y que hipótesis alternativa se está contrastando en la presente prueba. La hipótesis nula para el test ANOVA de un factor es que la media de todos los grupos o niveles es la misma, y la hipótesis alternativa es que al menos la media de uno de los grupos es diferente, es decir, que no todas las medias son iguales.

H_0: la media de todos los grupos es la misma ($\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$)

H_A: al menos la media de uno de los grupos es diferente (no todas las medias son iguales)

Recordar que no se trata de la media de la muestra, sino la media de la población a lo que se refieren las pruebas sobre las hipótesis.

La prueba ANOVA de un factor calcula un nivel de significancia (p-valor), que es la probabilidad de que las medias de las muestras de todos los grupos analizados sean al menos tan diferentes como las analizadas en el estudio, dando por válida la hipótesis nula. Si la probabilidad es suficientemente pequeña (asumiremos $p < 0{,}05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con lo que se puede concluir que no todas las medias de los diferentes grupos considerados son iguales, sino que existe al menos una que es diferente al resto. De forma alternativa, se rechaza la hipótesis alternativa y por lo tanto no se puede rechazar la hipótesis nula si la probabilidad es alta (asumiremos $p > 0{,}05$), concluyendo que las medias de todos los grupos son iguales. Cabe remarcar que no se acepta la hipótesis nula, sino que se afirma que ésta no puede ser rechazada. En el ejemplo que se muestra, $p = 0{,}000$, pero esto no se debe interpretar como que nivel de significancia es realmente cero, sino que el nivel de significancia es $p < 0{,}0005$. Por lo tanto, como $p < 0{,}0005 < 0{,}05$, se puede concluir que existe una diferencia significativa en la variable independiente "Resistencia Compresión (MPa)" en alguno de los grupos analizados, ya sea para 3, 7 o 28 días (aún no podemos determinar de cual se trata). En este punto cabe recordar que la prueba ANOVA de un factor determina si las medias de las poblaciones son iguales, y no las medias de las muestras elegidas y analizadas.

Dependiendo de si el resultado de la prueba ANOVA de un factor nos ha determinado si existe diferencia entre las medias de al menos uno de los grupos o no, se puede optar por dos caminos diferentes:

1- Si el resultado de la prueba ANOVA de un factor ha determinado que la media de la variable dependiente es igual para todos los grupos, se ha terminado el análisis.

2- Si por el contrario, se ha determinado que al menos uno de los grupos tiene una media de la variable dependiente diferente al resto de los grupos, se pasará a investigar donde radican exactamente esas diferencias.

La prueba a posteriori de Tukey es una buena prueba si se desea comparar todas las posibles combinaciones de diferencias de grupos cuando la hipótesis de homogeneidad de varianzas no ha sido violada. Esta prueba proporciona intervalos de confianza para la diferencia de medias entre los diferentes grupos y si esas diferencias de medias resultan estadísticamente significativas. La prueba se presenta resumida en una tabla como la siguiente:

Comparaciones múltiples

Variable dependiente: Resistencia Compresión (MPa)

	(I) Edad (días)	(J) Edad (días)	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95% Límite inferior	Límite superior
HSD de Tukey	3d	7d	-6,56327	1,98459	,003	-11,2371	-1,8894
		28d	-14,28308	1,98459	,000	-18,9569	-9,6092
	7d	3d	6,56327	1,98459	,003	1,8894	11,2371
		28d	-7,71981	1,98459	,000	-12,3937	-3,0460
	28d	3d	14,28308	1,98459	,000	9,6092	18,9569
		7d	7,71981	1,98459	,000	3,0460	12,3937

Figura 16. Ejemplo de resultado obtenido al ejecutar la prueba a posteriori de Tukey tras asumir igualdad de varianzas y que al menos uno de los grupos tiene una media diferente al resto.

Con tres grupos diferentes (3, 7 y 28 días) existen tres posibles combinaciones de diferencias entre grupos pero en la tabla anterior se muestras seis combinaciones. Esto es porque los datos se encuentran repetidos para cada combinación de grupos; es decir, 3 días vs 7 días, y la misma comparación pero girada, 7 días vs 3 días. Puede entonces obviarse a la hora de hacer el análisis la mitad de las filas y fijarse únicamente en aquellas en que estemos comparando grupos distintos.

La prueba a posteriori de Tukey toma como hipótesis nula que la diferencia de medias de los dos grupos analizados es cero y como hipótesis alternativa que esa diferencia de medias no es cero:

H_0: la diferencia entre las medias de las poblaciones de los 2 grupos es cero ($\mu 1 - \mu 2 = 0$)

H_A: la diferencia entre las medias de las poblaciones de los 2 grupos no es cero ($\mu 1 - \mu 2 \neq 0$)

La prueba a posteriori de Tukey calcula un nivel de significancia (p-valor), que es la probabilidad de que las medias de las muestras de los grupos analizados sean al menos tan diferentes como las analizadas en el estudio, dando por válida la hipótesis nula. Si la probabilidad es suficientemente pequeña (asumiremos $p < 0{,}05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con lo que se puede concluir que las medias de los dos grupos analizados son diferentes. De forma alternativa, se rechaza la hipótesis alternativa y por lo tanto no se puede rechazar la hipótesis nula si la probabilidad es alta (asumiremos $p > 0{,}05$), concluyendo que las medias de los dos grupos seleccionados son iguales. Recordar que no se acepta la hipótesis nula, sino que se afirma que ésta no puede ser rechazada. Es importante en este punto recordar que las hipótesis de la prueba hacen referencia a la media de las poblaciones, y no a las medias de las muestras analizadas. En el ejemplo, si se decide comparar la variable dependiente a 3 y a 7 días, se obtiene, $p = 0{,}003$ (es decir, $p < 0{,}05$). Por lo tanto, se puede concluir que la variable independiente "Resistencia Compresión (MPa)" es estadísticamente diferente para ambos grupos de población, es decir, se obtienen resultados diferentes para 3 y para 7 días. O expresado de otra manera, la diferencia de medias en la resistencia a compresión entre las muestras de analizadas a 3 y a 7 días resulta ser estadísticamente significativa. Lo que esto significa es que hay una probabilidad del 0,3% de obtener una diferencia de medias al menos tan grande como el obtenido si la hipótesis nula es verdadera (la hipótesis nula que indica que no hay ninguna diferencia entre la media de ambos grupos).

- La hipótesis de igualdad de varianzas es falsa. Cuando los resultados muestran que la hipótesis de homogeneidad de varianzas ha sido violada, se optará por realizar una prueba ANOVA robusta de un factor, como es la prueba ANOVA de un factor de Welch (es vez de la prueba ANOVA de un factor).

Pruebas robustas de igualdad de las medias

Resistencia Compresión (MPa)

	Estadístico[a]	gl1	gl2	Sig.
Welch	25,777	2	205,171	,000

a. Distribuidos en F asintóticamente.

Figura 17. Ejemplo de resultado obtenido al ejecutar la prueba ANOVA robusta de un factor de Welch asumiendo la no igualdad de varianzas.

Esta prueba se interpreta de la misma manera que una prueba ANOVA regular de un factor. Las hipótesis que se contrastan en la prueba ANOVA de Welch son las mismas que en la prueba ANOVA regular. La hipótesis nula para la prueba ANOVA de un factor de Welch es que la media de todos los grupos o niveles es la misma, y la hipótesis alternativa es que al menos la media de uno de los grupos es diferente, es decir, que no todas las medias son iguales.

H_0: la media de todos los grupos es la misma ($\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$)

H_A: al menos la media de uno de los grupos es diferente (no todas las medias son iguales)

Si la prueba ANOVA de Welch resulta estadísticamente significativa (nivel de significancia: $p < 0{,}05$), se puede concluir que no todas las medias de los grupos son iguales. Es decir, al menos un valor medio de uno de los grupos es diferente (ya que se ha rechazado la hipótesis nula). El p-valor en este ejemplo parece ser cero, pero no debe interpretarse así, sino que esto significa que $p < 0{,}0005$. Como $p < 0{,}05$ (ya que $p < 0{,}0005 < 0{,}05$), se puede concluir en este ejemplo que, al menos, la media de uno de los grupos se diferencia de la media de los demás grupos.

Dependiendo de si el resultado de la prueba ANOVA de un factor de Welch nos ha determinado si existe diferencia entre las medias de al menos uno de los grupos o no, se puede optar por dos caminos diferentes:

1- Si el resultado de la prueba ANOVA de Welch ha determinado que la media de la variable dependiente es igual para todos los grupos, se ha terminado el análisis.
2- Si por el contrario, se ha determinado que al menos uno de los grupos tiene una media de la variable dependiente diferente al resto de los grupos, se pasará a investigar donde radican exactamente esas diferencias.

La prueba a posteriori de Games-Howell es una buena prueba si se desea comparar todas las posibles combinaciones de diferencias de grupos cuando la hipótesis de homogeneidad de varianzas ha sido violada. Esta prueba proporciona intervalos de confianza para la diferencia de medias entre los diferentes grupos y si esas diferencias de medias resultan estadísticamente significativas. La prueba se presenta resumida en una tabla como la siguiente:

Comparaciones múltiples

Variable dependiente: Resistencia Compresión (MPa)

	(I) Edad (días)	(J) Edad (días)	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95% Límite inferior	Límite superior
Games-Howell	3d	7d	-6,56327	1,90434	,002	-11,0592	-2,0673
		28d	-14,28308	1,98784	,000	-18,9768	-9,5894
	7d	3d	6,56327	1,90434	,002	2,0673	11,0592
		28d	-7,71981	2,05859	,001	-12,5799	-2,8597
	28d	3d	14,28308	1,98784	,000	9,5894	18,9768
		7d	7,71981	2,05859	,001	2,8597	12,5799

Figura 18. Ejemplo de resultado obtenido al ejecutar la prueba a posteriori de Games-Howell tras asumir la no igualdad de varianzas y que al menos uno de los grupos tiene una media diferente al resto.

A la hora de interpretar los resultados y de determinar si existe una diferencia significativa entre cada uno de los grupos se procede de la misma manera que cuando se asumen varianzas iguales (las hipótesis que se contrastan en la prueba son las mismas), la única diferencia es que cuando no se han asumido varianzas iguales, se realiza la prueba a posteriori de Games-Howell en vez de la prueba a posteriori de Tukey.

Por último, como los grupos de la variable independiente no se encuentran en una escala continua, se procederá a la representación de los resultados mediante un gráfico de barras. En el gráfico de barras, se introducen las barras de error, que especifican los intervalos de confianza al 95%.

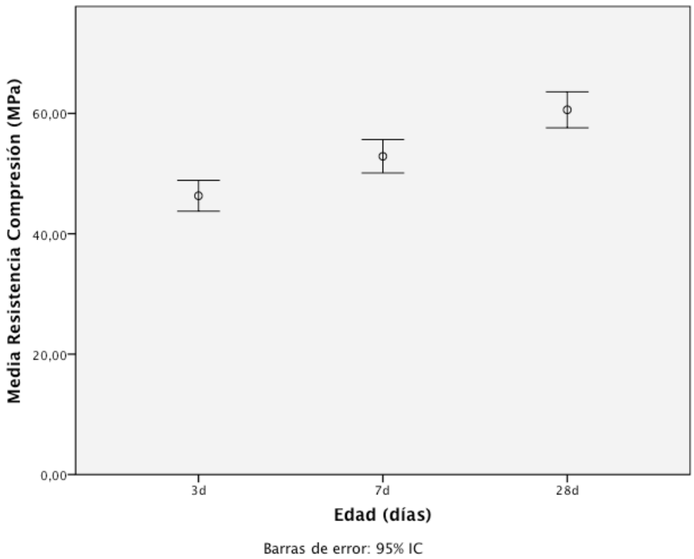

Figura 19. Ejemplo de gráfico de barras con intervalos de confianza del 95%.