

T. 4 – Inferencia estadística acerca de la relación entre variables (II): el análisis de varianza (ANOVA).

1. El ANOVA de un factor para muestras independientes

2. El ANOVA de un factor para muestras relacionadas

- En este capítulo se aborda el contraste de hipótesis acerca de la relación entre una variable categórica y una cuantitativa, ahora sin la restricción planteada en el capítulo previo de que la variable categórica sea dicotómica. El fundamento que subyace a la realización de este tipo de contraste sigue siendo el mismo: analizar si se pueden considerar estadísticamente significativas las diferencias entre las medias en la variable cuantitativa para los diferentes subgrupos definidos por la variable categórica. Ahora, al poder ser politómica la variable categórica, la comparación puede pasar a ser de dos a varias medias, siendo la prueba de significación tradicionalmente aplicada en este caso el conocido como análisis de varianza (ANOVA) (Fisher, 1935).

• A modo de síntesis de las pruebas de significación orientadas al contraste de hipótesis acerca de la relación entre una variable categórica y una cuantitativa, se puede plantear esta clasificación de las mismas atendiendo a los dos siguientes criterios:

1. El número de modalidades de la variable categórica: dicotómica versus politómica.
2. El modo en que estén constituidos los subgrupos definidos por la variable categórica:
 - 2.1. Los subgrupos están constituidos por distintos casos (muestras independientes);
 - 2.2. Los subgrupos están constituidos por distintos casos pero emparejados entre sí en los distintos subgrupos (muestras relacionadas), esto es, cuando los subgrupos han sido equiparados de acuerdo a terceras variables. Esto se suele hacer respecto a variables consideradas variables extrañas, como un medio de controlar el posible efecto de las mismas sobre las dos variables objeto de interés en el análisis.

El caso más extremo de diseño de investigación con muestras relacionadas es aquel en que los diferentes subgrupos están constituidos por los mismos sujetos –es el conocido como diseño de medidas repetidas, pues los mismos sujetos son medidos en diferentes ocasiones en una misma variable

	Muestras independientes	Muestras relacionadas
Variable categórica dicotómica (2 grupos o muestras → 2 medias)	<i>Prueba T para muestras independientes</i>	<i>Prueba T para muestras relacionadas</i>
Variable categórica politómica (3 o más grupos → 3 o más medias)	<i>ANOVA de un factor para muestras independientes</i>	<i>ANOVA de un factor para muestras relacionadas</i>

- Cuando se presupone la existencia de una relación asimétrica entre las variables, las diferentes pruebas de significación para contrastar la hipótesis de relación entre una variable categórica y una variable cuantitativa asumen que la primera sea la variable explicativa (independiente o predictora), mientras que la segunda, la variable de respuesta (dependiente o criterio).
- El ANOVA puede ser unifactorial (one-way) cuando únicamente hay un factor o variable independiente, o bien, factorial (multi-way ANOVA) cuando hay más de un factor o variable independiente.
- El ANOVA para muestras independientes recibe también el nombre de ANOVA intersujetos, mientras que el ANOVA para muestras relacionadas se denomina también ANOVA intrasujetos o de medidas repetidas.

1. El ANOVA de un factor para muestras independientes

- A continuación vamos a centrarnos en la prueba de significación estadística del Análisis de la varianza (ANOVA) comenzando por los pasos para a cabo el contraste de hipótesis acerca de la relación entre dos variables categóricas.

- Pasos en la realización de un ANOVA:

1. Se decide el nivel de riesgo (α) que se desea asumir en el contraste de hipótesis y se plantean las hipótesis de estudio (H_e) y nula (H_o).

Siendo J el número de modalidades de la variable categórica (o subpoblaciones que queremos comparar), la hipótesis nula sería que las medias en dichas poblaciones son iguales, esto es, que todas las muestras proceden de una misma población.

$$H_o : \mu_1 = \mu_2 = \mu_3 = \dots \mu_j$$

La hipótesis de estudio sería que, al menos, dos de las medias comparadas proceden de poblaciones diferentes.

2. Se calcula el estadístico de contraste. El estadístico de contraste es la razón F que se obtiene como el cociente entre las denominadas “Media cuadrática entregrupos” (MC entre) y “Media cuadrática intragrupos, residual o de error” (MC error):

$$F = \frac{MC \text{ entre}}{MC \text{ error}}$$

El estadístico F proviene de la denominada descomposición de la varianza de la variable cuantitativa en dos partes: la varianza explicada (aquella atribuible a la diferencia entre los grupos, es decir, a la variable categórica, también denominada entregrupos, o simplemente, varianza entre) y la varianza no explicada (también denominada varianza intragrupos o intra, residual o de error). El ANOVA consiste en obtener estimaciones de ambas varianzas, mediante el cálculo de las MC entre y MC error, y compararlas entre sí a través de un cociente. A medida que la MC entre supere a la MC error el valor del estadístico F será mayor y, por tanto, también lo será la evidencia en contra de la hipótesis nula.

El cálculo de todos los elementos conducentes a la obtención de la razón F se suele sintetizar en la denominada Tabla resumen del ANOVA:

Fuentes de variación	Sumas de cuadrados (SC)	Grados de libertad (gl)	Medias Cuadráticas (MC)	F	Sig
Entre	SC entre	J-1	MC entre = SC entre / gl entre	MC entre/ MC error	
Error	SC error	n-J	MC error = SC error / gl error		
Total	SC Total	n-1			

donde n es el número de sujetos y J el número de grupos

3. Se obtiene en la distribución F de Snedecor, con los grados de libertad entre (o gl del numerador) y los grados de libertad error (o gl del denominador), la probabilidad de obtener un valor del estadístico F como el obtenido o más extremo, esto es, el nivel de significación (Sig).

Se puede obtener este valor de probabilidad con el programa *Excel* si introducimos, en una casilla cualquiera de la hoja de cálculo, la siguiente fórmula con los valores entre paréntesis que nos interese:

$$=DISTR.F(F;gl_{entre};gl_{error})$$

En el caso del ANOVA todos los contrastes son unilaterales dado que la distribución F es asimétrica positiva y el valor de Sig corresponderá siempre a un área en la cola derecha de la distribución.

4. Decisión: se mantiene la H_0 si $Sig \geq \alpha$; por contra, se rechaza si $Sig < \alpha$.

Ejemplo: Sea el caso de una investigación que tiene por objeto analizar la relación existente entre la dosis de alcohol ingerida (1:Alta; 2:Media; 3:Nula) y el tiempo de reacción (ms). Se recogieron los siguientes datos de una muestra de 27 sujetos que participaron en el estudio. Supóngase que los resultados obtenidos fueron:

	N	Media	Desviación típica
Alto	9	55,89	16,586
Medio	9	28,78	21,141
Nulo	9	22,00	13,426
Total	27	35,56	22,366

En este ejemplo, la variable cuantitativa es la puntuación obtenida en el tiempo de reacción, mientras que la variable categórica es la dosis de alcohol ingerida. A continuación se muestran los pasos del ANOVA para contrastar si existen diferencias estadísticamente significativas entre las medias de los tres grupos a nivel poblacional:

1. Nivel de riesgo (α) = 0,05

$$H_0: \mu(\text{alta}) = \mu(\text{media}) = \mu(\text{nula})$$

$$H_e: \text{hay diferencias e.s. al menos entre un par de grupos}$$

Las tres medias muestrales parecen apoyar, en un principio, la hipótesis de que el tiempo de reacción difiere en función de la dosis de alcohol consumida.

2. Cálculo del estadístico de contraste F :

SPSS: Analizar > Comparar medias > ANOVA de un factor

ANOVA

Tiempo reacción					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	5788,222	2	2894,111	9,622	,001
Intra-grupos	7218,444	24	300,769		
Total	13006,667	26			

3. La distribución muestral del estadístico F para este ejemplo es la distribución F de Snedecor con 2 y 24 grados de libertad (gl entre e intra, respectivamente). El valor de significación, esto es, la probabilidad de obtener un valor igual o superior al obtenido en la muestra (9,622) siendo cierta la hipótesis nula, es igual a 0,001.

4. Decisión: $0,001 < 0,05$, por tanto, se rechaza la H_0 y se puede concluir, con un nivel de riesgo del 5%, que hay diferencias estadísticamente significativas al menos entre dos de las medias comparadas o, dicho de otra manera, que la dosis de alcohol sí que influye en el tiempo de reacción.

- La correcta aplicación de la prueba de significación estadística anterior requiere del cumplimiento de una serie de supuestos, entre ellos, uno de gran importancia es el conocido como supuesto de homogeneidad de las varianzas, esto es, las varianzas de la variable cuantitativa en las poblaciones comparadas deben ser iguales. El cumplimiento de este supuesto es evaluado habitualmente mediante la *prueba de Levene*, prueba de significación en la que se contrasta la hipótesis nula:

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

En caso de que no se satisfaga este supuesto (es decir, cuando el resultado de la prueba de Levene conduce al rechazo de la H_0), se aconseja realizar alguna de las pruebas basadas en la modificación del estadístico F (véase, por ejemplo, Pardo y San Martín, 2010). SPSS ofrece dos alternativas (la prueba de Welch y la de Brown y Forsythe).

Ejemplo sobre “tiempo de reacción” y “dosis de alcohol”. La hipótesis nula que se contrasta el realizar la prueba de Levene es:

$$H_0 : \sigma_{alta}^2 = \sigma_{media}^2 = \sigma_{nula}^2$$

La salida de SPSS al ejecutar el ANOVA incluye el resultado de la prueba de Levene. En este ejemplo se obtiene:

Prueba de homogeneidad de varianzas

Tiempo reacción

Estadístico de Levene	gl1	gl2	Sig.
,935	2	24	,406

Dado el nivel de significación obtenido para la prueba de Levene ($Sig = 0,406$), y asumiendo un nivel de riesgo $\alpha = 0,05$, se mantiene la hipótesis nula de igualdad de varianzas en la población, por lo que podemos considerar como satisfecho el cumplimiento de este supuesto. En caso contrario (si el nivel de significación de la prueba de Levene fuese inferior a $0,05$), se rechazaría la hipótesis nula de igualdad de varianzas y se debería realizar algunas de las pruebas alternativas al ANOVA mencionadas anteriormente. Para el ejemplo que nos ocupa el resultado de estas pruebas obtenido con SPSS sería:

Pruebas robustas de igualdad de las medias

Tiempo reacción

	Estadístico ^a	gl1	gl2	Sig.
Welch	11,223	2	15,507	,001
Brown-Forsythe	9,622	2	21,151	,001

a. Distribuidos en F asintóticamente.

- El ANOVA nos informa acerca de la significación estadística de la diferencia entre las medias y, en consecuencia, acerca de si la relación entre la variable categórica y la variable cuantitativa es estadísticamente significativa; sin embargo, no resulta obvio interpretar a partir del valor de F y su

significación cuál es la magnitud de la relación o tamaño del efecto. Además, al igual que ocurría con las pruebas de significación estudiadas en los temas anteriores, el valor de *Sig* obtenido está influido por el tamaño de la muestra. Uno de los indicadores del tamaño del efecto más utilizados al realizar un ANOVA es el denominado eta-cuadrado (η^2) y se obtiene como:

$$\eta^2 = \text{SC entre} / \text{SC total}$$

Este índice oscila entre 0 (independencia entre las variables) y 1 (relación perfecta entre las variables) e indica la variabilidad compartida por ambas variables y, más concretamente, el porcentaje de variabilidad de la variable de respuesta (cuantitativa) explicado por la variable categórica. Un problema asociado a este índice es que sobreestima dicha varianza y, por ello, se han propuesto distintas correcciones a este índice (véase Pardo y San Martín, 2010), siendo una de ellas la η^2 corregida:

$$\eta^2 \text{ corregida: } 1 - [(1 - \eta^2)(N - 1) / (N - J)]$$

Ejemplo sobre “tiempo de reacción” y “dosis de alcohol”. Los resultados son:

$$\eta^2 = 5788.22 / 13006.67 = 0,445$$

$$\eta^2 \text{ corregida: } 1 - [(1 - 0,445)(27 - 1) / (27 - 3)] = 0,399$$

Así, por tanto, se concluye que un 39,9% de la variabilidad observada en el tiempo de reacción se explica a partir de la dosis de alcohol ingerida.

- El rechazo de la *H₀* al realizar una ANOVA permite concluir que hay diferencias estadísticamente significativas al menos entre dos de las medias comparadas, pero esta conclusión global es poco precisa. Para especificar cuáles de los grupos comparados difieren significativamente, tras obtener un estadístico *F* significativo es necesario realizar las denominadas Comparaciones múltiples a posteriori (o post hoc). Existen distintas pruebas desarrolladas para ello de manera que al aplicarlas se controle la tasa de error tipo I fijada a priori (véase Pardo y San Martín, 2010). Cuando el objetivo es comparar todos los posibles pares de medias las pruebas más recomendadas son la HSD (honestly significant difference) de Tukey (si se cumple supuesto de que las varianzas poblacionales son iguales), o bien, la de Games-Howell (si se incumple dicho supuesto).

Ejemplo sobre “tiempo de reacción” y “dosis de alcohol”. Dado que se cumple el supuesto de homogeneidad de varianzas se debería obtener la prueba HSD de Tukey. El resultado de esta prueba obtenido con SPSS es:

Descriptivos

Tiempo reacción									
	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo	
					Límite inferior	Límite superior			
Alto	9	55,89	16,586	5,529	43,14	68,64	35	80	
Medio	9	28,78	21,141	7,047	12,53	45,03	6	70	
Nulo	9	22,00	13,426	4,475	11,68	32,32	4	47	
Total	27	35,56	22,366	4,304	26,71	44,40	4	80	

Comparaciones múltiples

Variable dependiente: Tiempo reacción

	(I) Alcohol	(J) Alcohol	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
HSD de Tukey	Alto	Medio	27,111*	8,175	,008	6,69	47,53
		Nulo	33,889*	8,175	,001	13,47	54,31
	Medio	Alto	-27,111*	8,175	,008	-47,53	-6,69
		Nulo	6,778	8,175	,689	-13,64	27,19
	Nulo	Alto	-33,889*	8,175	,001	-54,31	-13,47
		Medio	-6,778	8,175	,689	-27,19	13,64

El resultado obtenido permite concluir que hay diferencias estadísticamente significativas entre las medias del tiempo de reacción bajo las dosis alta y media ($P = 0,008$) y también bajo los niveles de dosis alta y nula ($P = 0,001$) siendo, en ambos casos, el tiempo de reacción significativamente superior bajo la dosis alta. Por el contrario, no hay diferencias estadísticamente significativas entre las medias del tiempo de reacción bajo los niveles de dosis media y nula ($P = 0,689$). En la tabla anterior también se presentan los $IC(95\%)$ de la diferencia de medias y, tal y como se expuso en el tema dedicado a la comparación entre dos medias, estos IC también permiten concluir sobre si la diferencia entre los pares de medias comparados es o no estadísticamente significativa en función de que el IC correspondiente incluya o no el valor 0 en su recorrido.

- En el informe técnico o de investigación deberían redactarse los resultados del ANOVA de alguna de estas formas equivalentes:

...hay un efecto estadísticamente significativo de la dosis de alcohol sobre el tiempo de reacción, $F(2,24) = 9,622, P = 0,001$.

...hay diferencias estadísticamente significativas en el tiempo de reacción en función de la dosis de alcohol ingerida, $F(2,24) = 9,622, P = 0,001$.

En caso de que el estadístico F sea estadísticamente significativo la conclusión anterior se debe complementar incluyendo el valor del tamaño del efecto y el resultado de las comparaciones múltiples a posteriori. Así, en el ejemplo:

... ..hay un efecto estadísticamente significativo de la dosis de alcohol sobre el tiempo de reacción, $F(2,24) = 9,622, P = 0,001, \eta^2$ corregida = 0,399. Las comparaciones múltiples a posteriori evidencian que hay diferencias estadísticamente significativas en el tiempo de reacción bajo las dosis de alcohol alta y media, $P = 0,008, IC95\% [6,69, 47,53]$ y bajo las dosis alta y nula, $P = 0,001, IC95\% [13,47, 54,31]$. En ambos casos el tiempo de reacción es significativamente superior bajo la dosis alta.

- La prueba t de Student estudiada en el tema anterior de comparación de 2 medias en muestras independientes puede considerarse un caso particular del ANOVA, de manera que si se eleva al cuadrado el valor de estadístico t se obtendrá el valor del estadístico F al realizar un ANOVA, y el valor de *Sig* será el mismo al realizar ambas pruebas.

Ejemplo: En el tema anterior se analizaron los datos procedentes del experimento sobre “choque emocional” y “recuerdo”. Los resultados al aplicar la prueba t para muestras independientes fueron:

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
Recuerdo	Se han asumido varianzas iguales	,400	,545	3,381	8	,010	4,000	1,183	1,271	6,729
	No se han asumido varianzas iguales			3,381	7,396	,011	4,000	1,183	1,232	6,768

Si los mismos datos son analizados mediante un ANOVA con SPSS se obtiene:

Prueba de homogeneidad de varianzas

Recuerdo

Estadístico de Levene	gl1	gl2	Sig.
,400	1	8	,545

ANOVA de un factor

Recuerdo

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	40,000	1	40,000	11,429	,010
Intra-grupos	28,000	8	3,500		
Total	68,000	9			

Se puede comprobar que $3,381^2 = 11,43$, y en ambas pruebas el valor $Sig = 0,010$. Además el resultado de la prueba de Levene para comprobar el supuesto de homogeneidad de las varianzas también es idéntico.

Por último, en caso de incumplimiento del supuesto del supuesto de homogeneidad de varianzas, también serían equivalentes las modificaciones propuestas para los estadísticos t y F . En el ejemplo, el resultado de las modificaciones del estadístico F obtenido con SPSS es:

Pruebas robustas de igualdad de las medias

Recuerdo

	Estadístico ^a	gl1	gl2	Sig.
Welch	11,429	1	7,396	,011
Brown-Forsythe	11,429	1	7,396	,011

a. Distribuidos en F asintóticamente.

- En conclusión, si se desea comparar dos medias se puede aplicar indistintamente la prueba t o el ANOVA, mientras que si se desea comparar 3 o más medias únicamente es aplicable el ANOVA.

2. El ANOVA de un factor para muestras relacionadas

- Esta prueba estadística recibe también el nombre de ANOVA intrasujetos o de medidas repetidas y su objetivo es el contraste de hipótesis acerca de la diferencia entre las medias de una misma variable medida en los mismos sujetos en diferentes momentos temporales, o bien, bajo diferentes condiciones experimentales. También este análisis es el adecuado cuando se desea comparar la media de distintas variables, siempre y cuando las variables tengan la misma escala de medida. Por ejemplo, cuando en un estudio de encuestas se pide a un grupo de sujetos que expresen su grado de satisfacción (en una escala de 0 a 10) con diferentes medidas para fomentar la participación ciudadana es adecuado comparar posteriormente las medias obtenidas en dichas variables para valorar cuáles son las medidas más o menos populares.
- Nótese que, por lo que se refiere a la estructura de los datos, en el caso de muestras relacionadas, las distintas columnas del archivo de datos corresponden a las variables cuantitativas medidas y no aparecerá ninguna variable categórica en dicho archivo.

- Pasos en la realización de un ANOVA:

1. Se decide el nivel de riesgo (α) que se desea asumir en el contraste de hipótesis y se plantean las hipótesis estadística y nula.

Siendo J el número de medias que se desea comparar, la hipótesis nula es que las medias son iguales.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots \mu_j$$

La hipótesis de estudio sería que, al menos, dos de las medias comparadas son diferentes.

2. Se calcula el estadístico de contraste. El estadístico de contraste es la razón F que se obtiene como el cociente entre las denominadas “Media cuadrática entregrupos” (MC entre) y la “Media cuadrática del error” (MC error):

$$F = \frac{MC \text{ entre}}{MC \text{ error}}$$

A diferencia del ANOVA para muestras independientes en que la varianza total se descomponía únicamente en dos partes (entregrupos y error), en este caso la varianza total se descompone en 3 partes: la varianza entregrupos, la varianza entresujetos y la varianza del error. Así, en el

ANOVA para muestras relacionadas, es posible calcular la varianza entresujetos (la que recoge las diferencias individuales) y extraerla de la varianza del error.

El cálculo de todos los elementos conducentes a la obtención de la razón F se suele sintetizar en la denominada Tabla resumen del ANOVA:

Fuentes de variación	Sumas de cuadrados (SC)	Grados de libertad (gl)	Medias Cuadráticas (MC)	F	Sig
Entregrupos	SC entregrupos	J-1	MC entre = SC entre / gl entre	MC entre/ MC error	
Entresujetos	SC entresujetos	n-1			
Error	SC error	(J-1) (n-1)	MC error = SC error / gl error		
Total	SC Total	n·J - 1			

donde n es el número de sujetos y J el número de grupos

- Se obtiene en la distribución F de Snedecor, con los grados de libertad entre (o gl del numerador) y los grados de libertad del error (o gl del denominador), la probabilidad de obtener un valor del estadístico F como el obtenido o más extremo, esto es, el nivel de significación (Sig). Este valor de probabilidad se puede obtener en Excel al igual que para el ANOVA con muestras independientes.
- Decisión: se mantiene la H_0 si $Sig \geq \alpha$; por contra, se rechaza si $Sig < \alpha$.

Ejemplo basado en una investigación sobre la relación entre la procedencia del vino y la calidad percibida del mismo (adaptado de Portell y cols., 2003, p. 31). En un estudio se quiso evaluar si la información que aparece en la etiqueta del vino acerca de su procedencia influye sobre la calidad percibida del mismo en su cata (valorada en una escala de 0 a 10). Para llevar a cabo este estudio, una misma muestra de 20 sujetos estimó la calidad de un vino en 3 ocasiones distintas. En cada una de ellas se les presentó para su cata una botella de vino que en realidad sólo cambiaba en su etiquetado (catalán, francés e italiano). Para cada sujeto se aleatorizó el orden de presentación de los 3 tipos de etiqueta.

En este ejemplo, la matriz de datos está compuesta por 20 filas y 3 columnas que corresponden a las 3 variables cuantitativas (la percepción de los sujetos bajo las 3 etiquetas). Las medias obtenidas en cada una de las condiciones fueron:

Estadísticos descriptivos

	N	Media	Desv. típ.
Catalán	20	7,25	1,020
Francés	20	6,90	1,165
Italiano	20	6,00	,973
N válido (según lista)	20		

A continuación se muestran los pasos del ANOVA para contrastar si existen diferencias estadísticamente significativas entre las medias de los tres grupos a nivel poblacional:

1. Nivel de riesgo (α) = 0,05

$$H_0: \mu(\text{catalán}) = \mu(\text{francés}) = \mu(\text{italiano})$$

$$H_e: \text{hay diferencias e.s. al menos entre un par de medias}$$

Las tres medias muestrales parecen apoyar, en un principio, la hipótesis de que la percepción del vino difiere en función de la etiqueta de procedencia.

2. Cálculo del estadístico de contraste F :

Fuentes de variación	Sumas de cuadrados (SC)	Grados de libertad (gl)	Medias Cuadráticas (MC)	F	Sig
Entregrupos	16,63	2	8,32	10,29	0,0003
Entresujetos		19			
Error	30,7	38	0,81		
Total		59			

La salida de *SPSS* al ejecutar el ANOVA para muestras relacionadas con los datos del ejemplo anterior es la siguiente:

SPSS: Analizar > Modelo Lineal General > Medidas repetidas

Pruebas de efectos intra-sujetos.

Medida: MEASURE_1

Fuente		Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
etiqueta	Esfericidad asumida	16,633	2	8,317	10,294	,000
	Greenhouse-Geisser	16,633	1,734	9,593	10,294	,001
	Huynh-Feldt	16,633	1,893	8,788	10,294	,000
	Límite-inferior	16,633	1,000	16,633	10,294	,005
Error(etiqueta)	Esfericidad asumida	30,700	38	,808		
	Greenhouse-Geisser	30,700	32,945	,932		
	Huynh-Feldt	30,700	35,960	,854		
	Límite-inferior	30,700	19,000	1,616		

En esta tabla se puede observar que los resultados obtenidos anteriormente son los que aparecen bajo las filas “Etiqueta” y “Error (etiqueta)” y “Esfericidad asumida” (éste es un supuesto del ANOVA para muestras relacionadas que se tratará posteriormente).

- La distribución muestral del estadístico F para este ejemplo es la distribución F de Snedecor con 2 y 38 grados de libertad (gl entre y gl error, respectivamente). El valor de significación, esto es, la probabilidad de obtener un valor igual o superior al obtenido en la muestra (10,3) siendo cierta la hipótesis nula es igual a 0,0003.
- Decisión: $0,0003 < 0,05$, por tanto, se rechaza la H_0 y se puede concluir, con un nivel de riesgo del 5%, que hay diferencias estadísticamente significativas al menos entre dos de las medias comparadas o, dicho de otra manera, que la etiqueta de procedencia sí influye en la percepción de la calidad del vino. No obstante, esta conclusión hay que precisarla analizando entré qué pares de etiquetas se dan las diferencias, mediante comparaciones múltiples a posteriori, tal y como se verá posteriormente.

- La correcta aplicación del ANOVA para muestras relacionadas requiere del cumplimiento de una serie de supuestos, entre ellos, uno de gran importancia es el conocido como supuesto de esfericidad, esto es, que sean homogéneas o iguales a nivel poblacional las varianzas de las variables resultantes de obtener las diferencias entre pares de puntuaciones. Lógicamente, este supuesto sólo es aplicable si comparamos 3 o más medias. En caso de incumplimiento de este supuesto, la razón F está sobreestimada, incrementándose la probabilidad de cometer un error de tipo I. El cumplimiento de este supuesto es evaluado habitualmente mediante la *prueba de Mauchly*, prueba de significación en la que se contrasta la H_0 de igualdad de las varianzas de las diferencias, por lo que si se mantiene la H_0 , ello significa que se satisface el supuesto de esfericidad.

- Cuando no se cumple el supuesto de esfericidad, existen diversas correcciones del ANOVA que modifican los grados de libertad del numerador y del denominador al multiplicarlos por un factor corrector denominado *Epsilon* y, en consecuencia, es también modificado el nivel de significación asociado a la razón *F* (Pardo y San Martín, 2010). SPSS calcula los tres siguientes factores correctores: el de Greenhouse-Geisser, el de Huynh-Feldt y el del límite inferior (éste último es el más conservador). Podemos optar por cualquiera de ellas (la de Greenhouse-Geisser se suele utilizar más en la práctica) y, por tanto, fijarnos en las filas correspondientes de la tabla.

Ejemplo sobre “percepción del vino”. La hipótesis nula que se contrasta al realizar la prueba de Mauchly es que las varianzas de las puntuaciones de diferencia son iguales, esto es, que se cumple el supuesto de esfericidad. El resultado al realizar esta prueba con SPSS es:

Prueba de esfericidad de Mauchly

Medida: MEASURE_1

Efecto intra-sujetos	W de Mauchly	Chi-cuadrado aprox.	gl	Significación	Epsilon ^a		
					Greenhouse e-Geisser	Huynh-Feldt	Límite-inferior
etiqueta	,847	2,999	2	,223	,867	,946	,500

Contrasta la hipótesis nula de que la matriz de covarianza error de las variables dependientes transformadas es proporcional a una matriz identidad.

- Puede usarse para corregir los grados de libertad en las pruebas de significación promediadas. Las pruebas corregidas se muestran en la tabla Pruebas de los efectos inter-sujetos.
- Diseño: Intercept
Diseño intra sujetos: etiqueta

En conclusión, dado el nivel de significación obtenido para la prueba de Mauchly es ($Sig = 0,223$) y asumiendo un nivel de riesgo $\alpha = 0,05$, se mantiene la hipótesis nula y, por tanto, se cumple el supuesto de esfericidad. En caso contrario (si el nivel de significación de la prueba de Mauchly fuese inferior a $0,05$), se rechazaría la hipótesis nula de esfericidad y se deberían interpretar los resultados del ANOVA bajo alguna de las correcciones propuestas.

- Al igual que en el ANOVA para muestras independientes, si la hipótesis nula de igualdad de medias es rechazada, se debe cuantificar el tamaño del efecto mediante el estadístico eta-cuadrado (η^2).

Ejemplo sobre “percepción del vino”. Este índice lo ofrece SPSS en la tabla de resultados del ANOVA bajo la columna Eta al cuadrado parcial:

Pruebas de efectos intra-sujetos.

Medida: MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial
etiqueta	Esfericidad asumida	16,633	2	8,317	10,294	,000	,351
	Greenhouse-Geisser	16,633	1,734	9,593	10,294	,001	,351
	Huynh-Feldt	16,633	1,893	8,788	10,294	,000	,351
	Límite-inferior	16,633	1,000	16,633	10,294	,005	,351
Error(etiqueta)	Esfericidad asumida	30,700	38	,808			
	Greenhouse-Geisser	30,700	32,945	,932			
	Huynh-Feldt	30,700	35,960	,854			
	Límite-inferior	30,700	19,000	1,616			

Así, por tanto, se concluye que un 35,1% de la variabilidad observada en la percepción del vino se explica por las diferentes etiquetas presentadas.

- El rechazo de la H_0 al realizar el ANOVA requiere, al igual que el caso de muestras independientes, evaluar entre qué pares de medias se dan las diferencias. Para ello es necesario realizar Comparaciones múltiples a posteriori. En este caso las comparaciones por pares se realizan aplicando la prueba t de Student para muestras relacionadas, pero controlando mediante la tasa de error tipo I fijada a priori mediante el método de Bonferroni (véase Pardo y San Martín, 2010).

Ejemplo sobre “percepción del vino”. El resultado obtenido con SPSS al realizar las comparaciones múltiples a posteriori es:

Factores intra-sujetos

Medida: MEASURE_1

etiqueta	Variable dependiente
1	Catalán
2	Francés
3	Italiano

Estadísticos descriptivos

	N	Media	Desv. típ.
Catalán	20	7,25	1,020
Francés	20	6,90	1,165
Italiano	20	6,00	,973
N válido (según lista)	20		

Comparaciones por pares

Medida: MEASURE_1

(I) etiqueta	(J) etiqueta	Diferencia entre medias (I-J)	Error típ.	Significación ^a	Intervalo de confianza al 95 % para diferencia ^a	
					Límite inferior	Límite superior
1	2	,350	,327	,892	-,508	1,208
	3	1,250*	,289	,001	,491	2,009
2	1	-,350	,327	,892	-1,208	,508
	3	,900*	,228	,003	,301	1,499
3	1	-1,250*	,289	,001	-2,009	-,491
	2	-,900*	,228	,003	-1,499	-,301

Basadas en las medias marginales estimadas.

*. La diferencia de las medias es significativa al nivel ,05.

a. Ajuste para comparaciones múltiples: Bonferroni.

El resultado obtenido permite concluir que hay diferencias estadísticamente significativas entre las medias de percepción del vino entre las etiquetas catalán e italiano ($P = 0,001$) y entre las etiquetas francés e italiano ($P = 0,003$) siendo en ambos casos la media más baja bajo la condición “italiano”. En la tabla anterior se presentan los IC(95%) de la diferencia de medias que también permiten concluir sobre si la diferencia entre los pares de medias comparadas es o no estadísticamente significativa.

- En el informe técnico o de investigación deberían redactarse los resultados del ANOVA de alguna de estas formas equivalentes:

...hay un efecto estadísticamente significativo de la etiqueta acerca de la procedencia del vino sobre la calidad percibida del mismo, $F(2,38) = 10,29, P < 0,0005$.

...hay diferencias estadísticamente significativas en la percepción del vino en función de la etiqueta sobre la procedencia del mismo, $F(2,38) = 10,29, P < 0,0005$.

En caso de que el estadístico F sea estadísticamente significativo la conclusión anterior se debe complementar incluyendo el valor del tamaño del efecto y el resultado de las comparaciones múltiples a posteriori. Así, para el ejemplo sobre “percepción del vino”:

... ..hay un efecto estadísticamente significativo de la etiqueta acerca de la procedencia del vino sobre la calidad percibida del mismo, $F(2,38) = 10,29, P < 0,0005, \eta^2 = 0,351$. Las comparaciones múltiples a posteriori evidencian que hay diferencias estadísticamente significativas en la percepción del vino entre las etiquetas catalán-italiano, $P = 0,001$, IC95% [0,491, 2,009] y francés-italiano, $P = 0,003$, IC95% [0,301, 1,499]. En ambos casos el vino bajo la etiqueta “italiano” es el peor valorado.

- Al igual que se mostró en el apartado sobre el ANOVA para muestras independientes, la prueba t de Student estudiada en el tema anterior de comparación de 2 medias en muestras relacionadas puede considerarse un caso particular del ANOVA, de manera que si se eleva al cuadrado el valor de estadístico t se obtendrá el valor del estadístico F obtenido al realizar un ANOVA, y el valor de Sig será el mismo para ambos estadísticos.

Ejemplo: En el tema anterior se analizaron los datos procedentes de 6 sujetos en la variable “Ansiedad” recogidos antes y después de una intervención. Los resultados al aplicar la prueba t para muestras relacionadas fueron:

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	pretest	8,0000	6	1,41421	,57735
	postest	6,0000	6	1,41421	,57735

Prueba de muestras relacionadas

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 pretest- postest	2,00000	1,26491	,51640	,67256	3,32744	3,873	5	,012

Si los mismos datos son analizados mediante un ANOVA se obtiene:

Pruebas de efectos intra-sujetos.

Medida: MEASURE_1

Origen		Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Tiempo	Esfericidad asumida	12,000	1	12,000	15,000	,012
	Greenhouse-Geisser	12,000	1,000	12,000	15,000	,012
	Huynh-Feldt	12,000	1,000	12,000	15,000	,012
	Límite-inferior	12,000	1,000	12,000	15,000	,012
Error(Tiempo)	Esfericidad asumida	4,000	5	,800		
	Greenhouse-Geisser	4,000	5,000	,800		
	Huynh-Feldt	4,000	5,000	,800		
	Límite-inferior	4,000	5,000	,800		

Se puede comprobar que $3,873^2 = 15$ y en ambas pruebas el valor $Sig = 0,012$. Cabe señalar que, en este caso no es aplicable el supuesto de esfericidad, dado que únicamente se comparan dos medias.

- En conclusión, si se desea comparar dos medias en muestras relacionadas se puede aplicar indistintamente la prueba t o el ANOVA para muestras relacionadas obteniéndose resultados equivalentes, mientras que si se desea comparar 3 o más medias únicamente es aplicable el ANOVA para muestras relacionadas.

Referencias

Fisher, R. A. (1935). *Design of experiments* (5ª ed.). Edinburgh: Oliver and Boyd.

Losilla, J. M., Navarro, B., Palmer, A., Rodrigo, M. F., y Ato, M. (2005). *Del contraste de hipótesis al modelado estadístico*. Tarrasa: CBS (www.edicionsapeticio.com).

Pardo, A., y San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. Madrid: Síntesis.