09

PSICOLOGÍA EXPERIMENTAL

Alexis Hancevich Leiva
Psicólogo Interno Residente.
Hospital Universitario General de Villalba.
Máster Universitario en Investigación en Psicología.

Verónica Ventero Portelas
Psicóloga Especialista en Psicología Clínica.
FEA Psicología Clínica Complejo Hospitalario de Toledo.

Juan Antequera Iglesias
Psicólogo Especialista en Psicología Clínica.
FEA Psicología Clínica Hospital Virgen de la Misericordia de Toledo.

Laura Hernangómez Criado
Doctora en Psicología.
Psicóloga Especialista en Psicología Clínica.
Psicoterapeuta acreditada por ASEPCO.
FEA Psicología Clínica Hospital Complejo Hospitalario de Toledo.
Hospital Universitario de Toledo. SESCAM.

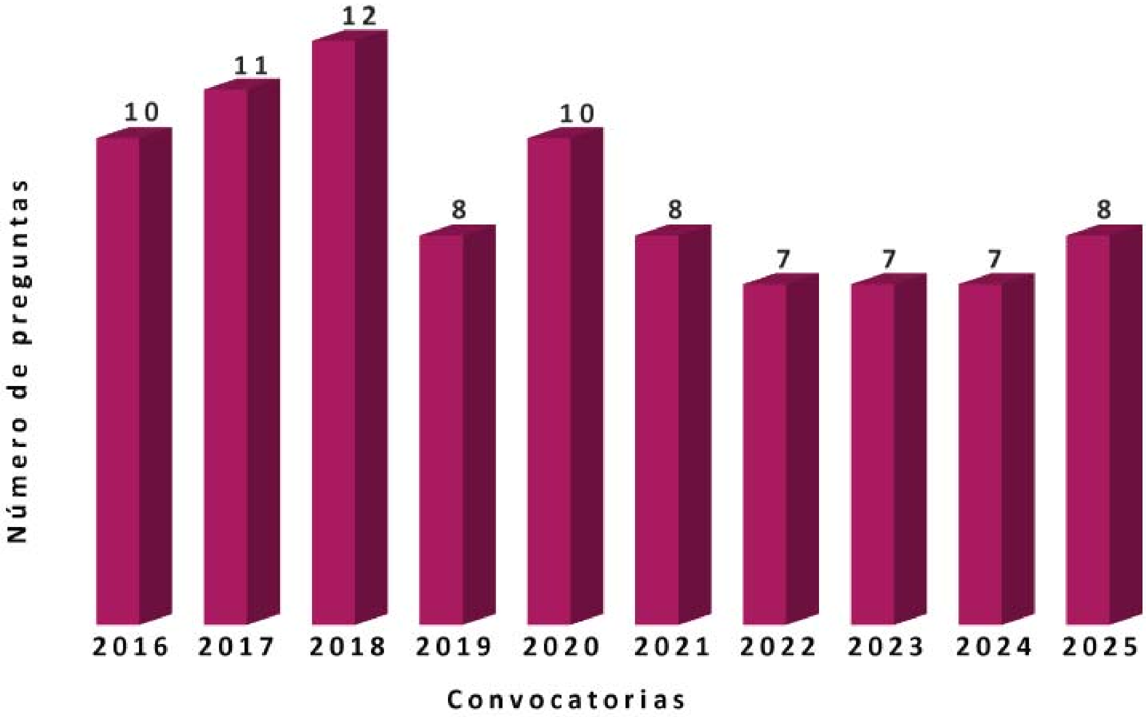
Carla Tejero Berzosa
Psicóloga Especialista en Psicología Clínica.
FEA Psicología Clínica Hospital Universitario del Tajo.

Rosa Ruiz Girón
Residente de Psicología Clínica en el Hospital Universitario de La Paz. Madrid.

MANUAL DE PSICOLOGÍA EXPERIMENTAL
ISBN obra completa: 978-84-10149-36-6
ISBN: 978-84-10149-48-9
Depósito Legal: M-7116-2026
EDICIÓN: Febrero 2026

Diseño de la cubierta: Rosa Ruiz Girón
Impresión: FRAGMA S.L.

EVOLUCIÓN DEL NÚMERO DE PREGUNTAS POR CONVOCATORIA Y TEMA



Temas	Convocatorias	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	
09 01 01 EL MÉTODO CIENTÍFICO: CARACTERÍSTICAS Y CLASIFICACIÓN		0	0	0	0	0	0	0	0	0	0	0
09 01 02 FUNCIÓN DEL DISEÑO, INTERVENCIÓN SOBRE LAS VARIABLES Y VALIDEZ		2	1	2	0	2	0	0	0	0	1	8
09 01 03 DISEÑOS EXPERIMENTALES		1	1	1	1	0	2	0	1	0	0	7
09 01 04 DISEÑOS CUASI-EXPERIMENTALES Y DISEÑOS EX POST FACTO		1	0	1	1	0	0	0	0	0	0	3
09 01 05 DISEÑOS N = 1 Y METODOLOGÍA DE ENCUESTAS		0	0	0	0	0	0	0	0	0	0	0
09 01 06 LA OBSERVACIÓN Y LA INVESTIGACIÓN CUALITATIVA		0	0	1	0	0	0	0	0	0	0	1
09 02 01 INTRODUCCIÓN A LA ESTADÍSTICA EN PSICOLOGÍA		0	1	0	0	0	0	0	0	0	0	1
09 02 02 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE UNA SOLA VARIABLE		1	0	0	0	2	0	1	1	1	0	6
09 02 03 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE DOS VARIABLES		0	1	2	1	0	2	1	1	0	0	8
09 02 04 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE TRES VARIABLES		0	0	0	0	0	1	1	0	0	0	2
09 02 05 PROBABILIDAD		0	0	0	0	0	0	0	0	0	0	0
09 02 06 FUNDAMENTOS BÁSICOS DE LA ESTADÍSTICA INFERENCIAL		1	2	0	1	1	0	0	2	2	0	9
09 02 07 TÉCNICAS NO PARAMÉTRICAS		1	0	0	0	0	0	0	1	1	0	3
09 02 08 PRUEBAS PARAMÉTRICAS		1	1	3	2	4	1	3	1	0	2	18
09 03 01 INTRODUCCIÓN A LA PSICOMETRÍA		0	0	0	0	0	0	1	0	0	0	1
09 03 02 TEORÍA CLÁSICA DE LOS TESTS		0	0	0	0	0	0	0	0	0	0	0
09 03 03 CÁLCULO DEL COEFICIENTE DE FIABILIDAD		1	1	2	1	0	0	0	0	2	1	8
09 03 04 VALIDEZ		0	1	0	0	0	1	0	0	0	2	4
09 03 05 TEORÍA DEL RASGO LATENTE		1	2	0	1	1	1	0	0	1	2	9
	Total de preguntas por convocatoria	10	11	12	8	10	8	7	7	7	8	

09 PSICOLOGÍA EXPERIMENTAL

09.01. FUNDAMENTOS TEÓRICOS Y DISEÑOS EXPERIMENTALES

09.02. ESTADÍSTICA

09.03. PSICOMETRÍA

ÍNDICE GENERAL

09.1. FUNDAMENTOS TEÓRICOS Y DISEÑOS EXPERIMENTALES

09 01 01. EL MÉTODO CIENTÍFICO: CARACTERÍSTICAS Y CLASIFICACIÓN

09 01 02. FUNCIÓN DEL DISEÑO, INTERVENCIÓN SOBRE LAS VARIABLES Y VALIDEZ

09 01 03. DISEÑOS EXPERIMENTALES

09 01 04. DISEÑOS CUASI-EXPERIMENTALES Y DISEÑOS EX POST FACTO

09 01 05. DISEÑOS N = 1 Y METODOLOGÍA DE ENCUESTAS

09 01 06. LA OBSERVACIÓN Y LA INVESTIGACIÓN CUALITATIVA

09.2. ESTADÍSTICA

09 02 01. INTRODUCCIÓN A LA ESTADÍSTICA EN PSICOLOGÍA

09 02 02. ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE UNA SOLA VARIABLE

09 02 03. ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE DOS VARIABLES

09 02 04. ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE TRES VARIABLES

09 02 05. PROBABILIDAD

09 02 06. FUNDAMENTOS BÁSICOS DE LA ESTADÍSTICA INFERENCIAL

09 02 07. TÉCNICAS NO PARAMÉTRICAS

09 02 08. PRUEBAS PARAMÉTRICAS

09.3. PSICOMETRÍA

09 03 01. INTRODUCCIÓN A LA PSICOMETRÍA

09 03 02. TEORÍA CLÁSICA DE LOS TESTS

09 03 03. CÁLCULO DEL COEFICIENTE DE FIABILIDAD

09 03 04. VALIDEZ

09 03 05. TEORÍA DEL RASGO LATENTE

ÍNDICE

09 02 08. PRUEBAS PARAMÉTRICAS

09 02 08. PRUEBAS PARAMÉTRICAS

ORIENTACIONES

En este tema se aborda de forma específica qué pruebas concretas utilizar según el contraste que queramos hacer. Todas las pruebas expuestas en este tema son pruebas paramétricas. De este tema lo más importante es conocer cuándo se utiliza cada prueba, así como los supuestos que hay que cumplir para poder utilizarlas.

ASPECTOS ESENCIALES

1. Para estimar el valor de una media en la población utilizaremos la prueba t en el caso de que la varianza sea desconocida, y la prueba z si la varianza es conocida. Para poder aplicar estas pruebas tendremos que cumplir independencia y normalidad.
2. Para estimar la diferencia entre dos medias utilizaremos la prueba t para diferencia de medias. Los supuestos que habrá que cumplir serán normalidad y homocedasticidad.
3. En el caso de tener más de dos medias, la alternativa será el análisis de la varianza (ANOVA). Si el análisis es intergrupo, tendremos que cumplir independencia, normalidad y homocedasticidad. Si el análisis es intragrupo tendremos que cumplir, además, esfericidad y aditividad.
4. Si queremos controlar estadísticamente una variable contaminadora aplicaremos un ANCOVA, que no es más que un ANOVA a la que se le añade una regresión lineal simple.

PREGUNTAS REPRESENTATIVAS

012. ¿Qué técnica se utiliza para examinar si la distribución de los datos se ajusta a una distribución teórica (como la distribución normal)?:

1) La prueba de Kolmogorov-Smirnov.
2) La prueba de Mauchly.
3) El test de rachas.
4) La prueba de Levene.

PIR 25, RC 1.

012. Una ventaja de utilizar la corrección de Bonferroni y el procedimiento HSD de Tukey y otras pruebas post hoc es que:

1) Aumenta la potencia estadística.
2) Disminuye la potencia estadística.
3) Disminuye la probabilidad de error tipo II.
4) Disminuye la probabilidad de error tipo I.

PIR 23, RC 4.

014. Seleccione la prueba estadística más adecuada para las siguientes dos hipótesis. H1: Las estimaciones de CI de los varones son más elevadas que las de las mujeres. H2: El CI medio de las mujeres es más elevado que el de los varones:

1) H1: Mann-Whitney; H2: t-test para muestras independientes.
2) H1: p (rho) de Spearman; H2: r de Pearson.
3) H1: t-test para muestras relacionadas; H2: Chi-cuadrado.
4) H1: T de Wilcoxon; H2: Prueba binomial de signos.

PIR 23, RC 1.

006. En un diseño multigrupos al azar, si el investigador desea realizar comparaciones a posteriori entre todos los pares de medias, en el caso de cumplimiento de los supuestos del modelo estadístico y para garantizar la mayor potencia y control del error de tipo I, la prueba más adecuada es:

1) El procedimiento de Bonferroni.
2) HSD de Tukey.
3) Prueba de Dunnet.
4) Prueba de Fisher.

PIR 22, RC 2.

007. En el ámbito de las medidas del tamaño del efecto, para evaluar el impacto de un determinado tratamiento, indique cuál de las siguientes afirmaciones es <u>CORRECTA</u>:

1) Eta cuadrado es equivalente al coeficiente de determinación de un modelo de regresión.
2) Omega cuadrado es un estimador sesgado por el tamaño muestral y se recomienda utilizar eta cuadrado en su lugar.
3) Delta de Cliff es un estimador paramétrico del tamaño del efecto.
4) Sólo la “d” de Cohen puede informar acerca del tamaño de efecto de un tratamiento.

PIR 22, RC 1.

032. Cuando en el análisis de varianza unifactorial no se cumple el supuesto de homocedasticidad, en vez de la F ordinaria, ¿cuál de las siguientes pruebas convendría emplear?:

1) Brown-Forsythe.

2) Shapiro Wilk.
3) Levene.
4) Mauchly.

PIR 21, RC 1.

001. Atendiendo a las características del diseño de covarianza, señale la afirmación INCORRECTA:

1) El análisis asociado a este diseño es el análisis de covarianza, el cual combina las ventajas del análisis de la varianza y el análisis de regresión.
2) Se trata de un diseño en el que se utiliza un procedimiento de control experimental sobre la variable perturbadora o covariable.
3) Es un diseño que permite reducir la varianza de error e incrementar la potencia estadística.
4) Este diseño se utiliza en la metodología cuasiexperimental con el propósito de controlar el sesgo de selección.

PIR 20, RC 2.

1. INTRODUCCIÓN

En este tema nos centraremos en los contrastes que más se suelen utilizar para sacar conclusiones estadísticas.

Existen dos grandes grupos de pruebas a la hora de hacer contrastes estadísticos:

1. Pruebas paramétricas: Todas estas pruebas sirven para estimar parámetros. Aunque tienen mucha potencia estadística y se obtiene mucha información con ellas, necesitaremos cumplir varios supuestos para poder aplicarlas. Para poder utilizar dichas pruebas tendremos que partir de muestras grandes (mayores de 30 sujetos por grupo), así como que la **variable dependiente esté definida a nivel de intervalo o razón** (PIR 08, 231). Además de estas dos condiciones generales para todas las pruebas paramétricas, tendrán algunos supuestos particulares que iremos analizando a lo largo del tema.

2. Pruebas no paramétricas: en general, este tipo de pruebas nos dan información acerca de elementos que no son parámetros (tipos de distribución, independencia…), aunque también tendremos versiones de pruebas no paramétricas que sirven para estimar parámetros. En este tipo de análisis nos centraremos en el próximo tema.

Como regla general, siempre que sea posible intentaremos aplicar una prueba paramétrica.

PARAMÉTRICAS	NO PARAMÉTRICAS (de distribución libre)
Hipótesis sobre parámetros	Hipótesis sobre bondad de ajuste, independencia, etc.
Nivel de medida cuantitativo	Nivel de medida bajo
Supuestos	Aplicables a muestras pequeñas
No pierden información	Robustez
Alta potencia estadística	

2. SUPUESTOS PRUEBAS PARAMÉTRICAS

La aplicación de pruebas paramétricas como ANOVA exige el cumplimiento de algunos supuestos, dependiendo de la prueba concreta éstos serán más o menos, en el caso del ANOVA que veremos después, nos exige el cumplimiento de supuestos como normalidad, independencia y homocedasticidad, y en ANOVA de medidas repetidas, además, debemos cumplir esfericidad y aditividad. Se exponen aquí todos los supuestos al inicio para que sean útiles al estudio, indicando posteriormente en cada prueba los que será necesario cumplir.

2.1. NORMALIDAD (PIR 17, 222; PIR 25, 12)

Las puntuaciones han de seguir una distribución normal (curva normal). Es el más flexible de todos los supuestos y su incumplimiento tiene poca incidencia por ejemplo, sobre el test F. Para verificar esta condición existen las pruebas de bondad de ajuste (como Chi cuadrado) y también otras pruebas específicas. Por ejemplo, se ejecutará el test de Shapiro-Wilk cuando el tamaño de la muestra es igual o menor a 50 unidades y la prueba de Kolmogorov-Smirnov cuando tenemos más de 50 sujetos. Esta última prueba examina si la distribución se ajusta a la curva normal con media y varianza conocidas. Cuando no se conocen los parámetros de la población (algo muy frecuente, con media y varianza desconocidas), se aplica una corrección a la prueba anterior llamada corrección de Lilliefors. En estadística, el Test de Shapiro-Wilk se usa para contrastar la normalidad de un conjunto de datos. Se plantea como hipótesis nula que una muestra $x_1, ..., x_n$ proviene de una población normalmente distribuida. La hipótesis nula se rechazará si *W es* demasiado pequeño, su valor puede oscilar entre 0 y 1. Siendo la hipótesis nula que la población está distribuida normalmente, si el p-valor es menor a alfa (nivel de significancia) entonces la hipótesis nula es rechazada (se concluye que los datos no vienen de una distribución normal). Si el p-valor es mayor a alfa, se concluye que no se puede rechazar dicha hipótesis; al igual que ocurre normalmente en cualquier contraste de hipótesis. Si se incumple el supuesto de normalidad de manera importante se puede recurrir a pruebas no paramétricas o realizar una transformación logarítmica de los datos para que se ajusten a la curva normal. Su incumplimiento no aumenta la probabilidad de cometer error tipo I.

2.2. HOMOCEDASTICIDAD

Hace referencia a la igualdad de las varianzas de las subpoblaciones de las que proceden los distintos grupos o conjuntos de datos **(PIR 12, 18)**. El supuesto de homogeneidad de las varianzas es un supuesto importante desde el punto de vista analítico, ya que su incumplimiento puede afectar tanto a la probabilidad de cometer un error de tipo I, como a la potencia de la prueba estadística. Hay diferentes opiniones entre los teóricos sobre la distorsión de F si se produce heterocedasticidad. En lo que la mayoría de autores parecen estar de acuerdo es en que, ante grupos no equilibrados, niveles moderados de heterogeneidad pueden hacer que el error de tipo I o el alfa real sea sustancialmente diferente del alfa nominal fijado a priori por el investigador.

Para comprobar si se cumple esta condición se aplica una prueba estadística que consiste en un contraste, donde la hipótesis nula es la igualdad de varianzas y la alternativa la desigualdad. El estadístico se distribuye según χ^2 y las pruebas más utilizadas son: Prueba de Bartlett, Prueba C de Cochran y Prueba de Levene. Se podría utilizar una prueba F para comparar la varianza de los grupos, pero el uso de F requiere que las distribuciones sean normales. El uso de test de Levene (es decir, los valores absolutos de las desviaciones de la media) es más robusto, y el uso de Brown-Forsythe (es decir, los valores absolutos de las desviaciones de la mediana) es aún más robusto. Lo que hacen del test de Levene y la prueba de Brown-Forsythe, pruebas más robustas, es que se realizan sobre la transformada de datos, mientras que la razón F de grupo de varianzas (Hartley de la prueba) utiliza los datos en bruto. La transformación de los datos en cuestión son los valores absolutos de las desviaciones (de la media, en el caso de Levene y de la mediana, en el caso de Brown-Forsythe).

El estadístico o prueba de Welch (también conocida como prueba T o W de Welch) es utilizada únicamente cuando se puede asumir que las dos varianzas poblacionales son diferentes (los tamaños muestrales pueden o no ser iguales) y por lo tanto deben ser estimadas por separado. El estadístico de Brown-Forsythe tiene un uso similar, aunque más ligado al ANOVA **(PIR 21, 32)**; la otra opción para en caso de incumplimiento de la homocedasticidad es realizar una transformación logarítmica de los datos o transformar el modelo mediante la técnica de los mínimos cuadrados **(PIR 06, 26)**. En general, para muchos autores, cuando los datos no cumplen el supuesto de homogeneidad de las varianzas cabe recurrir a varios procedimientos para corregir la heterocedasticidad, tales como, por ejemplo, el procedimiento de O'Brien, la prueba F conservadora, la F de Brown y Forsythe y la prueba T o W de Welch; tal y como se ha señalado previamente.

2.3. INDEPENDENCIA

Las puntuaciones han de ser independientes entre sí. El supuesto de independencia se garantiza si la muestra se ha obtenido mediante un muestreo aleatorio y los sujetos han sido asignados aleatoriamente a las condiciones experimentales. Por lo tanto, la selección aleatoria de los sujetos permite obtener observaciones mutuamente independientes y, por lo tanto, errores no correlacionados, asegurando el cumplimiento del supuesto de independencia. El supuesto es especialmente relevante si se pretende aplicar de forma correcta el análisis de la varianza, ya que su violación afecta considerablemente al nivel de significación y a la potencia de la prueba estadística. Algunos autores han demostrado que la dependencia entre las observaciones produce un serio incremento en la tasa de error de tipo I. Vallejo añade que la violación de la independencia también puede llevarnos a cometer un error tipo II. Para comprobar el supuesto de independencia se pueden utilizar test como el de Durbin-Watson, Box-Ljung o Rachas. El estadístico de Durbin-Watson, es una prueba que se utiliza para detectar la presencia de autocorrelación (una relación entre los valores separados el uno del otro por un intervalo de tiempo dado, por ejemplo) en los residuos (errores de predicción) de un análisis de la regresión **(PIR 18, 35)**. Cuando los mismos sujetos van a pasar por una serie de tratamientos experimentales, la variable dependiente, es siempre la misma (medidas repetidas). El supuesto de independencia de las observaciones es muy importante, en el caso de los diseños con medidas repetidas, dado que son los mismos sujetos los que reciben cada una de las condiciones experimentales, es muy probable que aparezca correlación entre sus puntuaciones y, por tanto, que se incumpla este supuesto (independencia de las observaciones). En el caso de incumplir este supuesto habría que transformar el modelo mediante la técnica de mínimos cuadrados **(PIR 16, 230)**.

2.4. ESFERICIDAD

Este supuesto solo se exige cuando los diseños son de medidas repetidas, también se denomina circularidad. En estos casos se requiere que las varianzas de las diferencias entre cada par de medias de medidas repetidas sean constantes. Siempre que se cumple este supuesto, el cociente de medias cuadráticas sigue la distribución F, a pesar de que exista covarianza entre las observaciones (no claro cumplimiento del supuesto de independencia). Una forma particular de esfericidad es la denominada simetría compuesta o combinada, cuya presencia exige que las correlaciones entre todos los pares de medidas repetidas sean iguales, el incumplimiento de la simetría combinada no implica el incumplimiento del supuesto de esfericidad pero indica una sospecha de transgresión de dicho supuesto. En los casos en los que no se cumple el supuesto de esfericidad, se obtienen estimaciones positivamente sesgadas de la razón F. Cuanto mayor sea el grado de covariación, mayor la probabilidad de que los resultados de la prueba F (ANOVA) resulten estadísticamente significativos, al "inflarse" el valor de F. Esto significa que aumenta la probabilidad de error tipo I. Existen varias pruebas para comprobar si se cumple este supuesto, la más utilizada es el Test W de Mauchly **(PIR 20, 6)**, aunque se ha de señalar que esta prueba es sensible al incumplimiento del supuesto de normalidad. Por otra parte, el supuesto de simetría combinada puede verificarse aplicando la prueba de Box.

En el caso de que no se cumpla el supuesto de esfericidad, se pueden adoptar tres alternativas:

- Simplificar los grados de libertad asociados al numerador (k – 1) y al denominador (n – 1) (k – 1) de la razón entre varianzas, de forma que la F sea mucho más conservadora.
- Corregir los grados de libertad de la razón F mediante el $\hat{\varepsilon}$ de Greenhouse y Geisser (multiplicar por dicho factor corrector).
- Aplicar el análisis multivariante de la varianza (MANOVA).

Las dos primeras alternativas son estrategias univariadas, mediante las que se llevan a cabo determinadas correcciones en el estadístico F, con el objetivo de reducir la probabilidad de cometer un error de tipo I en la decisión estadística. Ambas se engloban bajo la denominación de modelo mixto univariante. La tercera alternativa requiere el cumplimiento de menos supuestos en el análisis y se conoce como modelo multivariante. El MANOVA supone que cada una de las medidas que se realizan repetidamente al sujeto se considera como una variable dependiente, pudiendo aislar los errores puntuación por puntuación y controlarlos uno a uno. Algunos autores afirman que el modelo mixto es más potente que el modelo multivariante, siempre que se satisfaga el supuesto de esfericidad. Cuando este supuesto no se cumple, el modelo mixto tiene mayor potencia que el multivariante con muestras pequeñas de sujetos, mientras que el modelo multivariante es más potente, a medida que aumenta el número de unidades experimentales. El criterio de decisión entre ambas estrategias estaría en torno a un tamaño muestral de 20 + k sujetos, siendo k el número de tratamientos, otros autores hablan de un criterio de decisión más flexible cercano a un tamaño muestral de 10 + k sujetos.

2.5. ADITIVIDAD

Supone la no interacción sujeto x tratamiento. No obstante, es muy frecuente que se produzca una interacción entre la variable de tratamiento y la variable sujeto, lo que significa que existe una porción de la puntuación observada que no es función aditiva, ni del efecto principal del sujeto, ni del efecto principal del tratamiento. En este caso, el modelo matemático que representa más adecuadamente los datos es el modelo de no aditividad de los efectos. En consecuencia, el componente residual incluye tanto la variación debida al error como la forma específica en la que reacciona cada sujeto ante los diferentes tratamientos. La prueba de Tukey, permite saber si las desviaciones de las puntuaciones de los sujetos, con respecto a los promedios de los tratamientos, son diferentes de un sujeto a otro. Si los datos se ajustan al modelo de no aditividad, se estaría probablemente transgrediendo el supuesto de aditividad. Si sujetos y tratamientos interaccionan se dice que se ha violado el supuesto de aditividad; esta interacción se encuadra dentro de la varianza error, por lo que si esta aumenta, disminuirá el valor de F artificialmente, aumentando la probabilidad para mantener la Ho, aunque realmente sea falsa (error tipo II). Si se observa que la aditividad se incumple, puede utilizarse un procedimiento de transformación de las puntuaciones consistente en aplicar logaritmos a cada una de ellas.

RECUERDA

De cara al examen PIR es de vital importancia aprender qué pruebas sirven para comprobar qué supuestos: Kolmogorov Smirnov para normalidad, Levene para homocedasticidad...

3. ANÁLISIS DE REGRESIÓN Y CORRELACIÓN

Podemos poner a prueba si la ecuación de regresión que hemos obtenido en la muestra se podría generalizar a la población (análisis de la regresión) o si la correlación que aparece entre dos variables a nivel muestral es generalizable (análisis de la correlación).

Para poder aplicar el análisis de la regresión o de la correlación han de cumplirse una serie de condiciones previas como son: normalidad, independencia y homocedasticidad.

3.1. ANÁLISIS DE LA REGRESIÓN

Como ya vimos en el tema 3 se puede construir una recta de regresión con la intención de pronosticar las puntuaciones de un sujeto en una variable teniendo la puntuación de la otra variable.

Si recordamos, esta ecuación la llamábamos "ecuación de regresión lineal", y tenía la siguiente forma:

$$Y = bX + a$$

donde b y a están basados en la muestra a la que pertenecen.

Si quisiéramos extrapolar esto a la población, la fórmula sería:

$$Y = \beta X + \alpha$$

El parámetro que se suele estimar en este tipo de contrastes es β, es decir, la pendiente de la ecuación de regresión a nivel poblacional. El estadístico utilizado para poner a prueba las hipótesis acerca de β es **T**, y se distribuye según la t de Student, con n – 2 grados de libertad.

3.2. ANÁLISIS DE LA CORRELACIÓN

En algunas ocasiones nos interesará saber si la correlación que hemos obtenido a nivel muestral es significativa y generalizable a nivel poblacional. Para sacar conclusiones de este estilo tendremos que utilizar el análisis de la correlación.

Esto significa poner a prueba el parámetro ρ_{xy}. Utilizaremos para ello el estadístico **T** (con una fórmula distinta a la utilizada en el análisis de la regresión) con *n* – 2 grados de libertad. Como en el resto de los contrastes, si $p \leq \alpha$, rechazamos hipótesis nula, sea cual sea el valor de la correlación **(PIR 02, 165)**.

4. CONTRASTE DE HIPÓTESIS SOBRE UNA SOLA MEDIA

El contraste de hipótesis sobre una media sirve para tomar decisiones acerca del verdadero valor poblacional que corresponde a la media de una variable. Nos encontramos, por tanto, ante un diseño con una muestra.

Para hacer este tipo de contraste es necesario poner a prueba dos supuestos: normalidad **(PIR 05, 86)** e independencia de las puntuaciones, las cuales hemos explicado en el apartado anterior.

En el caso de conocer la varianza de la población que partimos, utilizaremos el estadístico **Z** para tomar decisiones acerca de la media. Este estadístico se distribuye según la normal, con media 0 y desviación 1.

Aunque lo más habitual es no conocer la varianza poblacional. En este caso, utilizaremos el estadístico **T** **(PIR 06, 30)**, que se distribuye según la t de student, con *n* – 1 grados de libertad.

5. CONTRASTE DE HIPÓTESIS SOBRE DOS MEDIAS

Quizá el contraste más frecuente sea el de saber si la diferencia entre las medias de dos grupos es significativa a nivel poblacional. Cuando, por ejemplo, se desea evaluar la eficacia de algún tratamiento o algún tipo de intervención, se seleccionan aleatoriamente dos grupos de sujetos; a uno de ellos se le aplica el tratamiento y a otro no; tras esto, se comparan las medias de los dos grupos en la variable de interés para determinar si difieren o no y, por tanto, si el tratamiento aplicado es o no eficaz.

Si se dan las condiciones apropiadas, el contraste de hipótesis sobre dos medias es el idóneo para comparar dos grupos de sujetos en alguna variable de interés **(PIR 06, 32)**.

Dentro de este tipo de contrastes, tendremos dos modalidades:

5.1. CONTRASTE DE HIPÓTESIS SOBRE DOS MEDIAS INDEPENDIENTES

Utilizaremos este tipo de contraste cuando hayamos utilizado dos grupos y los estemos comparando. Esto es, el ejemplo que acabamos de poner.

Al trabajar con dos medias independientes lo estamos haciendo con dos poblaciones distintas de las que extraemos, independientemente, dos muestras aleatorias de tamaños n_1 y n_2 (ej.: hombres y mujeres).

Para hacer este tipo de contrastes utilizaremos el estadístico **T** **(PIR 03, 42; PIR 12, 229)**, que se distribuye según t de Student con $n_1 + n_2 - 2$ grados de libertad.

Los supuestos que tendremos que cumplir para aplicar este estadístico son normalidad y homocedasticidad, aunque T es una prueba muy robusta (es decir, que incluso incumpliendo los supuestos el estadístico no varía mucho), por lo que se podrá utilizar con tamaños muestrales menores de 30 sujetos, o incluso incumpliendo alguno de los supuestos. Sólo tendremos que recurrir a pruebas alternativas en el caso de que se incumplan ambos supuestos y los tamaños muestrales sean pequeños **(PIR 00, 17)**.

5.2. CONTRASTE DE HIPÓTESIS SOBRE DOS MEDIAS RELACIONADAS

Hablamos de muestras relacionadas cuando un grupo de sujetos es evaluado dos veces. Si queremos comparar el rendimiento de un grupo de sujetos con problemas de aprendizaje en dos tareas de habilidad diferentes, podemos evaluar el rendimiento de cada sujeto en ambas tareas y comparar los promedios obtenidos en ellas; tendremos dos muestras de puntuaciones relacionadas, porque ambas pertenecen a los mismos sujetos.

Las medidas repetidas no son la única forma (aunque tal vez sí la más frecuente) de generar muestras relacionadas. También tenemos dos muestras relacionadas cuando, en lugar de medir a los mismos sujetos en dos ocasiones, utilizamos pares de sujetos. Por ejemplo, en un estudio sobre relaciones maritales podría interesarnos preguntar a los miembros de una muestra de matrimonios por su grado de satisfacción marital a fin de evaluar si los maridos se sienten, en promedio, más satisfechos o menos que sus esposas. Aquí, a cada individuo sólo le tomamos una medida, pero cada matrimonio, como una unidad, contribuye con un par de puntuaciones. Parece razonable asumir que si un miembro de la pareja se siente muy satisfecho con su matrimonio, el otro miembro de la pareja también se sentirá satisfecho, y viceversa; por lo que las puntuaciones de ambas muestras estarán relacionadas.

Para sacar conclusiones utilizaremos el estadístico **T** para medidas relacionadas, que se distribuye según la t de Student con $n-1$ grados de libertad.

Los supuestos a cumplir son normalidad y homocedasticidad, pero como hemos dicho en el punto 4.1 sólo se recurrirá a pruebas alternativas si se incumplen ambos criterios y la muestra es pequeña.

6. CONTRASTE DE HIPÓTESIS SOBRE MÁS DE DOS MEDIAS: ANÁLISIS DE VARIANZA UN SOLO CRITERIO DE CLASIFICACIÓN (ANOVA I)

Supongamos que queremos comparar la eficacia de tres técnicas terapéuticas distintas en el tratamiento del trastorno de pánico. Podemos, si fuera posible, reunir a todos los sujetos de la población de un distrito diagnosticados con este trastorno y repartirlos aleatoriamente en tres subgrupos o poblaciones. A continuación a cada una de las subpoblaciones le aplicaríamos un tipo de terapia. Al finalizar los tratamientos podríamos considerar como variable dependiente la frecuencia de ataques al mes. Como tenemos tres subpoblaciones, calcularíamos la media de ataques en cada una, obteniendo tres medias poblacionales:

(μ_1, μ_2, μ_3). Comparando las medias veríamos qué técnica ha sido más eficaz en la reducción de ataques de pánico. No obstante, esto resultaría imposible si las subpoblaciones fueran muy elevadas. En cambio, lo que sí es más factible es elegir aleatoriamente de cada subpoblación una muestra (de tamaños n_1, n_2 y n_3), y a continuación a la primera muestra le aplicaríamos la técnica terapéutica 1, a la segunda la técnica 2 y a la tercera la técnica 3.

Tras aplicar los tres tratamientos obtendríamos tres medias:

$$\overline{Y}_1, \overline{Y}_2 \text{ e } \overline{Y}_3$$

A partir de los resultados en las tres muestras, el análisis de varianza pretende hacer inferencias acerca de las medias (μ_1, μ_2, μ_3) de las tres subpoblaciones. Más en concreto, lo que se pretende es someter a comprobación experimental la hipótesis nula:

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

Obviamente, la hipótesis alternativa es la contraria, esto es, que existen diferencias entre las medias, a nivel poblacional, en función del tratamiento. Si se acepta la hipótesis nula, el investigador tiene que concluir que estadísticamente las tres técnicas son igual de efectivas; si se rechaza la nula, y en consecuencia se acepta la alternativa, el investigador puede concluir que no todos los tratamientos son igual de efectivos.

El análisis de varianza (ANOVA) siempre se va a referir a hipótesis acerca de medias y no de varianzas **(PIR 03, 53; PIR 10, 259)**. Su nombre se debe a que utiliza para la comprobación de hipótesis un análisis o descomposición de la variabilidad total de las observaciones de las muestras en dos componentes aditivos.

En este primer apartado del tema nos vamos a referir a un solo criterio de clasificación o factor (ANOVA I) (en nuestro ejemplo, el tipo de técnica terapéutica). Si el experimentador ha procedido bien, las puntuaciones de las tres muestras sólo van a diferir en función de la técnica empleada.

El análisis de varianza se va a complicar un poco más si además de un criterio de clasificación (tipo de técnica) empleáramos otro (nivel cultural). Si considerásemos tres niveles culturales obtendríamos (3 x 3) 9 subgrupos. En este caso el efecto podría deberse al tipo de técnica terapéutica, al nivel cultural o bien al efecto o interacción entre ambos. En el apartado siguiente presentaremos algunas nociones básicas del análisis de varianza con doble criterio de clasificación (ANOVA II).

6.1. CONCEPTOS BÁSICOS EN EL ANÁLISIS DE VARIANZA

Siguiendo con el ejemplo propuesto al inicio de este apartado, podemos concebir tres tipos de variabilidades en el conjunto de puntuaciones tras aplicar las tres técnicas terapéuticas:

Variabilidad total:

Si no nos detenemos a distinguir grupos, y reunimos todas las puntuaciones, vemos que lógicamente varían unas respecto a otras, por distintos motivos. Esta variabilidad total se puede atribuir principalmente a dos fuentes de variación: variabilidad intergrupo e intragrupo o error **(PIR 07, 255)**.

Variabilidad intergrupo:

Las puntuaciones del grupo 1 es probable que difieran respecto a las del 2 y, consiguientemente, respecto a las del 3, en función del distinto tratamiento que se está aplicando a cada grupo. O por lo menos, eso es lo que espera el experimentador, pues si sucede, podrá concluir que no todos los métodos son iguales.

Variabilidad intragrupo:

Por mucho que se haya esforzado el investigador en la construcción de cada uno de los grupos, le resultará imposible que todos y cada uno de los miembros de un grupo sean iguales entre sí.

Factores como la motivación, la inteligencia, el cansancio, las expectativas traídas al experimento, etc. van a funcionar como una potente fuente de variación sobre las puntuaciones. Este tipo de variabilidad también recibe el nombre de varianza error o residual, pues contribuye a contaminar los resultados del experimento.

Ahora bien, si la mayor parte de la variabilidad total se debe a la variabilidad intragrupo (por factores que nada tienen que ver con el criterio de clasificación), o lo que es lo mismo, si la variabilidad intragrupo es muy superior a la intergrupo, podremos concluir que el criterio de clasificación no es una fuente muy potente de variación sobre las puntuaciones, y por lo tanto que los diferentes tratamientos no crean variaciones significativas en las puntuaciones (en este caso en el número de ataques por mes).

Por el contrario, si la mayor parte de la variabilidad total se debe a la variabilidad intergrupo, el experimentador podrá concluir que son las diferencias entre las técnicas empleadas las responsables de las variaciones en el número de ataques de pánico al mes, y no otro tipo de factores (motivación, cansancio, etc.) asociados a la variabilidad error o intragrupo.

El criterio para decidir requiere de la construcción de un estadístico. Dicho estadístico de contraste es F, puesto que se obtiene a partir de un cociente entre varianzas (varianza intergrupo entre varianza intragrupo).

6.2. ESQUEMA DEL ANÁLISIS DE VARIANZA

Antes de proceder al contraste estadístico el investigador distribuye los datos de acuerdo a la siguiente tabla:

TERAPIA 1	TERAPIA 2	.....	TERAPIA k
Y_{11} Y_{21} ...... Y_{i1}	Y_{12} Y_{22} ...... Y_{i2}	....	Y_{1j} Y_{2j} ...... Y_{ij}
$\overline{Y}_1$	$\overline{Y}_2$	.....	$\overline{Y}_j$

El número de categorías o grupos que genera el criterio de clasificación se denomina (K). Genéricamente (Y_{ij}) representa la puntuación de un sujeto cualquiera perteneciente a un grupo cualquiera. Cada grupo constará de un determinado número de sujetos, siendo posible que difieran los grupos en cuanto al número de sujetos que contienen. Cuando el número coincide, se habla de modelo equilibrado o balanceado **(PIR 02, 170)**.

A continuación, con estos datos, el investigador calcula las siguientes expresiones:

$$\underbrace{\Sigma\Sigma(Y_{ij}-\overline{Y})^2}_{\text{SCt}} = \underbrace{\Sigma\, nj(\overline{Y}_j-\overline{Y})^2}_{\text{SCter}} + \underbrace{\Sigma\Sigma(Y_{ij}-\overline{Y}_j)^2}_{\text{SCtra}}$$

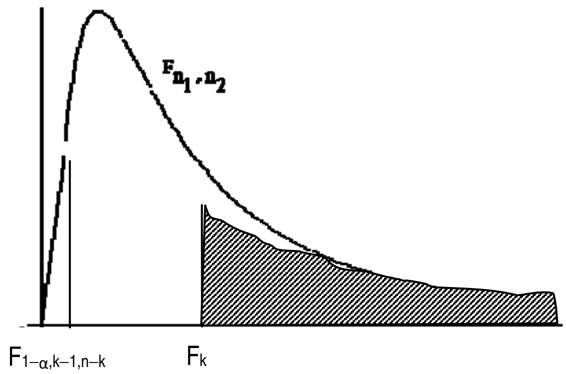
Suma cuadrática total SCt	Suma cuadrática intergrupo SCter	Suma cuadrática intragrupo SCtra

Si nos fijamos, las anteriores expresiones son los numeradores de la fórmula matemática de la varianza. Las sumas cuadráticas reflejan respectivamente, las variabilidades total, intergrupo e intragrupo.

FUENTE DE VARIACIÓN	SUMAS CUADRÁTICAS	GRADOS DE LIBERTAD	MEDIAS CUADRÁTICAS	ESTADÍSTICO DE CONTRASTE
INTER + INTRA	Scter + SCtra	k – 1 + n – k	Mcter = Scter / k – 1 + Mctra = Sctra / n – k	$F = \frac{MCter}{MCtra}$
TOTAL	SCt	n – 1		

Después de estos cálculos, el investigador puede hallar el estadístico de contraste F. En el caso de obtener un valor de F menor que la unidad, se deberá a que nuestros datos no se ajustan a los requisitos exigibles para realizar un ANOVA o bien a que el modelo utilizado no es adecuado. En estos casos, algunos autores, entre ellos Ostle, aconsejan que si tenemos una F con n_1 y n_2 grados de libertad, menor que la unidad, calculemos el valor 1/F y lo comparemos con el valor de la distribución de F con los grados de libertad cambiados, es decir n_2 y n_1. Si tras hacer esto el valor es menor que el de las tablas, podremos proceder a aceptar la hipótesis nula del Análisis de Varianza sin atribuirle mayor importancia al tema; en caso contrario (cuando el valor 1/F es mayor que el de las tablas) nos indicará que los supuestos no se están cumpliendo o que el modelo utilizado no es el adecuado, debiendo buscar un nuevo modelo que describa mejor los datos observados y el fenómeno bajo investigación. Por ello, según indica Pardo, una F significativa no necesariamente tiene que ir acompañada de la presencia de un resultado experimental relevante **(PIR 20, 2)**.

Si $F_k \geq F_{1-\alpha,\, k-1,\, n-k}$ entonces no puedo aceptar H_0, esto es, tengo que aceptar que existen diferencias a nivel poblacional entre los grupos (entre sus medias). F_k se refiere a la F empírica, es decir, a la que obtenemos a partir de la aplicación de las fórmulas, mientras que $F_{1-\alpha,\, k-1,\, n-k}$ hace referencia a la F teórica, es decir, a la que obtenemos en las tablas de distribución de densidad de probabilidad.



Si el estadístico de contraste nos permite rechazar la hipótesis nula, el siguiente paso es averiguar entre qué grupos se producen las diferencias significativas. Existen varios procedimientos para ver qué pares de medias difieren; uno de ellos es la prueba de **Scheffé**. Esta prueba se realiza sobre todos los pares de medias posibles:

$$\left|\bar{Y}_i - \bar{Y}_j\right| \Big/ \sqrt{MCtra\,(1/n_i + 1/n_j)} \geq \sqrt{(k-1)\,(F_{1-\alpha,\,k-1,\,n-k})}$$

$$A \geq B$$

Si A ≥ B entonces el par de medias (μ_i/μ_j) difieren entre sí.

Además de comprobar entre qué medias hay diferencias significativas, otra pregunta que puede plantearse es si las afirmaciones sobre las diferencias entre las medias son o no importantes. Un modo de medir esa importancia es averiguar qué parte de la variabilidad total puede ser atribuida a los tratamientos.

Existen diversos índices o pruebas utilizados como medida de la magnitud o tamaño del efecto experimental, es decir, que hacen referencia a la varianza explicada: eta-cuadrado (η^2), épsilon-cuadrado (ε^2) y omega-cuadrado (ω^2).

En cuanto a eta cuadrado (η^2), se indica la posibilidad de su utilización para calcular la magnitud del efecto del tratamiento, es decir, la proporción de varianza en la variable dependiente que es explicada por las diferencias entre los grupos; y se calcula por medio de la división de la suma cuadrática (SC) entre grupos y la SC total. Sin embargo, para diseños factoriales, el tamaño del efecto brinda un valor η^2 para cada efecto principal. La principal desventaja de eta cuadrado es que se trata de un estimador muy sesgado de la verdadera proporción de varianza explicada, de forma que su uso no es aconsejable por este motivo **(PIR 19, 160)**.

Épsilon-cuadrado (ε^2) sería una mejor estimación de la fuerza de la asociación en una población de lo que lo es η^2, aunque ε^2 presenta algunas dificultades en diseños factoriales y no queda claro si se puede generalizar su uso en los diferentes diseños experimentales (como si se puede hacer con otros índices como omega-cuadrado).

Finalmente, ω^2 (omega cuadrado) indica la proporción de varianza en la variable dependiente que es explicada por los niveles de la variable independiente y provee una medida de la fuerza relativa de una variable independiente. ω^2 no es afectado por muestras de tamaño pequeñas, por lo tanto omega-cuadrado es insensible a las variaciones en el tamaño de la muestra.

Para poder aplicar el análisis de varianza han de cumplirse una serie de condiciones previas, que ya han sido explicadas en este mismo tema: normalidad, independencia y homocedasticidad.

6.3. UN CASO PARTICULAR: EL ANÁLISIS DE VARIANZA CON MEDIDAS REPETIDAS

Hasta ahora planteábamos que k tratamientos eran aplicados a K grupos de personas, para, posteriormente, valorar los efectos diferenciales de los tratamientos mediante la comparación de las medias de cada grupo.

Ahora bien, es posible someter a un solo grupo de personas a k tratamientos secuencialmente: primero el tratamiento 1 y medimos su resultados en la variable dependiente, después el tratamiento 2 y medimos, ..., así hasta el tratamiento K.

En estos casos hemos eliminado una fuente de error, pues al tratarse del mismo grupo de personas no cabe la posibilidad de plantearse una fuente de variación debida a la desigualdad de los grupos.

El procedimiento a seguir es muy semejante al planteado con el ANOVA de medidas independientes. Las diferencias van a radicar en el tipo de fuentes de variabilidad. Cuando se trabaja con medidas repetidas, la **variabilidad total** se va a descomponer en:

Varianza intersujetos:

Es la variabilidad producida por las diferencias interindividuales de los sujetos que conforman la muestra. Se manifiesta por puntuaciones diferentes obtenidas por sujetos distintos dentro de un mismo tratamiento. Si los sujetos sometidos a los tratamientos fueran todos iguales, sus puntuaciones en la variable dependiente no diferirían entre sí, ya que el tratamiento es el mismo. La varianza intersujeto debe afectar por igual a todos los tratamientos, pues los sujetos, con sus diferencias interindividuales, son los mismos en todos ellos. No constituye formalmente una fuente de variables contaminadoras. La misma técnica estadística se encarga de segregar de la variabilidad total esta varianza, antes de proceder a examinar la influencia de los tratamientos.

Varianza intrasujetos:

Es la varianza que nos interesa, y que se define como las variaciones en las respuestas de los sujetos en la variable dependiente pero bajo influencia de tratamientos diferentes. Si los tratamientos influyen, las puntuaciones de los sujetos variarán de tratamiento a tratamiento. No obstante, la varianza intrasujetos se descompone en dos partes:

1. Varianza intertratamiento.
2. Varianza error o residual.

1. **Varianza intertratamiento:** Es la parte de la variación intrasujeto debida exclusivamente a la acción de los tratamientos. Es la varianza que interesa destacar. Para concluir que la variable independiente influye en la dependiente, ésta deberá resultar de mayor cuantía que la varianza error.

2. **Varianza error o residual:** Es aquella parte de la varianza intrasujeto debida a la influencia de variables extrañas insuficientemente controladas (efectos de la práctica o fatiga no corregidos suficientemente, falta de fiabilidad en el material empleado, etc.).

El experimentador quiere saber si la mayor parte de la variabilidad total es atribuible a los diferentes tratamientos o bien a la variabilidad residual. El estadístico de contraste es por tanto la proporción de variabilidad intertratamiento, expresado en términos de medias cuadráticas:

$$F = MCtra / MCres$$

Si $F \geq F_{(1 - \alpha)\ (k-1),\ (n-1)\ (k-1)}$ entonces rechazo la hipótesis nula, esto es, considero que no todos los tratamientos son igual de eficaces.

Por último, en los diseños de medidas repetidas o intragrupos, a la hora de aplicar una prueba estadística como es el ANOVA, es necesario comprobar dos supuestos adicionales a los tres habituales que se exigen en el análisis de varianza (independencia, normalidad, homocedasticidad), estos son esfericidad y aditividad.

7. CONTRASTE DE HIPÓTESIS SOBRE DIFERENCIA DE MÁS DE DOS MEDIAS: ANÁLISIS DE VARIANZA: DOBLE CRITERIO DE CLASIFICACIÓN (ANOVA II)

Cuando comenzamos este tema con el ejemplo de la eficacia de distintos tratamientos en la disminución de la frecuencia mensual de ataques de pánico, señalamos que además de comprobar la eficacia diferencial de las técnicas terapéuticas podíamos investigar cómo influía una segunda variable (nivel cultural), por ejemplo.

En estos casos nos encontramos con un ANOVA con dos criterios de clasificación. Estos dos criterios funcionan como dos variables independientes, cuyo efecto vamos a medir sobre una variable dependiente. Los criterios son también llamados factores y frecuentemente identificados con letras latinas mayúsculas (A y B) **(PIR 06, 22)**. Los factores o criterios no tienen por qué coincidir en el número de categorías o niveles. La multiplicación de sus categorías nos da el número de combinaciones de tratamientos o grupos posibles (2 × 3 = 6; 5 × 2 = 10, etc.).

No hay que olvidar que la influencia sobre la variable dependiente no es simple y directa: A puede no influir en la VD, B puede también no influir y sin embargo sí hacerlo la interacción entre A y B.

Aquí vamos a desarrollar brevemente el ANOVA para un doble criterio de clasificación, pero es posible aplicar esta técnica para tres y más criterios o factores (A × B × C; A × B × C × D; etc.).

Para realizar el ANOVA II se exigen las mismas condiciones previas que en el ANOVA I: **Independencia, Homocedasticidad y Normalidad**.

En el ANOVA II se procede de modo similar que en el ANOVA I, esto es, el investigador parte de un número determinado de subpoblaciones; en este caso tantas como combinaciones posibles entre las categorías de los dos criterios de clasificación (2 × 3 = 6; 5 × 2 = 10, etc.). Como no puede disponer de esas subpoblaciones, aplica el número de combinaciones resultante a un número igual de grupos de sujetos (Veremos en psicología experimental como no siempre esto es así, pues son posibles diseños de medidas repetidas).

A \ B	1	2	......	c	
1	μ_{11}	μ_{12}	......	μ_{1c}	$\mu_{1.}$
2	μ_{21}	μ_{22}	....	μ_{2c}	$\mu_{2.}$ ......
	μ_{11}	μ_{12}	......	μ_{fc}	$\mu_{f.}$
	$\mu_{.1}$	$\mu_{.2}$	......	$\mu_{.c}$	$\mu_{..}$

La tabla que hemos representado recoge las medias de las puntuaciones en todas y cada una de las subpoblaciones. Así, por ejemplo, la casilla que recoge la media subpoblacional (μ_{11}) alude a la media de la subpoblación que está sometida al nivel 1 del criterio de clasificación A y simultáneamente al nivel 1 del criterio de clasificación B. Las casillas marginales recogen las medias de las subpoblaciones que soportan los respectivos niveles de un criterio. Así, por ejemplo, la media ($\mu_{1.}$) corresponde a la subpoblación afectada por el nivel 1 del factor A. La media ($\mu_{.1}$) pertenece, en cambio a la subpoblación afectada por el nivel 1 del factor B.

El investigador construye una tabla idéntica a la anterior, sólo que en las casillas tanto conjuntas como marginales, figuran medias muestrales y no poblacionales.

No obstante, las hipótesis se formulan acerca de los parámetros. En el caso del ANOVA II, la hipótesis nula no es una, sino tres. Atendiendo a la tabla deducimos las siguientes hipótesis nulas:

1. Hipótesis nula (H_{01}) o efecto fila:

Esta hipótesis nula afirma que todos los niveles del factor A son igual de eficaces, esto es:

$$\mu_{1.} = \mu_{2.} = ... = \mu_{f.}$$

Se denomina efecto fila, porque las medias se obtienen de las respectivas filas, esto es, de los f niveles del criterio de clasificación A. Esta hipótesis nula se expresa también mediante la siguiente expresión:

$$\alpha_{r.} = 0$$

2. Hipótesis nula (H_{02}) o efecto columna:

Esta hipótesis hace referencia a la no existencia de diferencias entre los niveles del factor B, esto es:

$$\mu_{.1} = \mu_{.2} = ... = \mu_{.c}$$

Aparece frecuentemente bajo la siguiente nomenclatura:

$$\beta_{.S} = 0$$

3. Hipótesis nula (H_{03}) o efecto interacción:

Esta hipótesis afirma que no hay diferencias entre las fxc tipos de tratamientos. En definitiva significa que no existe interacción significativa entre el factor A y el B. Expresada matemáticamente, se refleja igualando las medias subpoblacionales de las casillas conjuntas:

$$\mu_{11} = \mu_{12} = ... = \mu_{1c} = \mu_{21} = \mu_{22} = ... = \mu_{2c} = \mu_{11} = \mu_{12} = ... = \mu_{fc}$$

La ausencia del efecto interacción se expresa también como:

$$\gamma_{rs} = 0$$

7.1. CONCEPTOS BÁSICOS EN EL ANOVA DE DOBLE CRITERIO

El ANOVA II trata también de comprobar las diferencias de medias mediante el análisis de la descomposición de la varianza. En este caso la variabilidad total de las puntuaciones se descompondría en cuatro tipos de variabilidad:

Variabilidad interfila o debida al factor A:

El factor A tiene determinados niveles (técnica terapéutica 1, técnica terapéutica 2, etc.), los cuales pueden ocasionar variaciones en las puntuaciones de la variable dependiente. Al menos, eso espera el experimentador.

Variabilidad intercolumna o debida al factor B:

El mismo caso pero referido al factor B. Parte de la variabilidad total de las puntuaciones es posible que sea atribuible a los distintos niveles del factor B.

Variabilidad interacción:

Parte de la variabilidad total de las puntuaciones pudiera no quedar explicada por el factor A o por el B, sino que se debería a la interacción entre ambos.

Variabilidad error:

Por último, parte o mucha de la variabilidad total de las puntuaciones podría deberse a las diferencias interindividuales existentes en cada uno de los grupos, que subsisten a pesar de que el experimentador intente que los sujetos sean similares.

Al comparar diferentes grupos, aumentamos la probabilidad de cometer Error Tipo I (a más comparaciones, más probabilidad de error). A la probabilidad de Error Tipo I, dependiendo del número de comparaciones, lo denominamos tasa de error por familia de comparaciones **(PIR 11, 249)**.

Si la proporción de V. Interfila es elevada en relación a la V. Error, podría concluirse que las variaciones en el factor A producen efectos diferenciales estadísticamente.

Si la proporción de V. Intercolumna es elevada en relación a la V. Error, podría concluirse que las variaciones en el factor B producen efectos diferenciales estadísticamente.

Si la proporción de V. Interacción es elevada en relación a la V. Error, podrá concluirse que las distintas combinaciones entre los niveles de ambos factores producen efectos diferenciales estadísticamente.

El criterio decisorio es, en este caso, no un estadístico de contraste, sino tres, que recogen las proporciones reseñadas. Si sus respectivos valores caen dentro de las regiones críticas, se rechazarán las respectivas hipótesis nulas.

Si en el ANOVA II resulta significativa la hipótesis de los factores por separado y de la interacción, es más fiable interpretar la interacción y no cada uno de los factores por separado. Por ejemplo, si realizamos un análisis de varianza con tres factores A, B y C y encontramos que los factores principales (A, B, C) son significativos, así como la interacción B x C (ninguna de las restantes interacciones son significativas), debemos analizar el factor principal A, así como la interacción B x C para interpretar los datos **(PIR 11, 248)**.

7.2. ESQUEMA DEL ANOVA II

El esquema de cálculo es similar al ANOVA I, salvo lo que ya hemos comentado, que existen cuatro fuentes de variabilidad y tres estadísticos de contraste, uno para cada hipótesis nula. Estos estadísticos se distribuyen también según Fisher.

FUENTE DE VARIACIÓN	SUMAS CUADRÁTICAS	GRADOS DE LIBERTAD	MEDIAS CUADRÁTICAS	ESTADÍSTICOS DE CONTRASTE
Interfila (A)	SCf	f – 1	MCf	F_{01} = MCf/MCe
Intercolumna (B)	SCc	c – 1	MCc	F_{02} = MCc/MCe
Interacción (A x B)	SCint.	(f – 1) (c – 1)	MCint.	F_{03} = MCint./MCe
Intragrupo (error)	SCerror	(n – 1) (fc)	MC error	
Total	SCt	nfc – 1		

No reproducimos aquí los cálculos de las SC por ser muy complejos y no aportar más información al esquema general del procedimiento.

Como en el ANOVA I, también es necesario, si alguna o todas las hipótesis nulas han sido rechazadas, comprobar entre qué niveles dentro de cada factor (en el caso de rechazar las hipótesis nulas 01 y 02) se producen las diferencias estadísticamente significativas. El mismo procedimiento se seguirá si se rechaza la hipótesis nula 03, para determinar entre qué combinaciones de niveles (de entre las cxf posibles) existen diferencias significativas, pues es frecuente que la interacción sólo se produzca entre determinados niveles de uno y otro factor y no entre el resto. Para estos contrastes se emplea también la prueba de Scheffé, o cualquier otra prueba a posteriori.

7.3. CONTRASTES O COMPARACIONES MÚLTIPLES

La razón F es un test ómnibus de comprobación de hipótesis o una prueba de significación general, es decir, prueba si la varianza explicada en un conjunto de datos es significativamente mayor que la no explicada. En consecuencia, el objetivo de dicha prueba consiste en verificar si las medias de los grupos de tratamiento, consideradas conjuntamente, presentan mayores diferencias de las que cabe esperar por azar. Esta información resulta muy útil para inferir si la variable independiente ejerce influencia sobre la conducta, pero no permite conocer la naturaleza de tal efecto.

Por ello, cuando la variable independiente es cualitativa y se obtiene una F significativa, resulta conveniente plantear hipótesis que realicen predicciones más específicas sobre los efectos de la VI. Dado que en este tipo de diseños hay más de dos grupos de tratamiento, existe más de una comparación posible, por ello pueden plantearse diversas hipótesis particulares susceptibles de contrastarse mediante los análisis denominados contrastes o comparaciones múltiples.

Las comparaciones múltiples entre las medias de los grupos pueden ser de dos tipos: comparaciones planificadas o a priori y comparaciones no planificadas o a posteriori.

Comparaciones planificadas o a priori

Las comparaciones planificadas son aquellas que se plantean antes de llevar a cabo el análisis de la varianza y obedecen a intereses que tiene el investigador en la fase previa a la realización del experimento.

A su vez, las planificadas se subdividen en dos categorías: contrastes no ortogonales y contrastes ortogonales o independientes entre sí.

Entre las pruebas planificadas o a priori encontramos la **Corrección de Bonferroni o Prueba Dunn-Bonferroni,** la cual nos permitiría llevar a cabo comparaciones múltiples entre pares de medias planificadas o a priori y está considerada como se considera el método más apropiado para controlar la tasa de error de tipo I por familia de comparaciones, cuando se pretende analizar un subconjunto de todas las posibles comparaciones, y éstas están formuladas a priori. Encontramos también la

Prueba de Scheffé, la cual se realiza sobre todos los pares de medias posibles y puede utilizarse tanto para comparaciones a priori como a posteriori.

La principal desventaja de las comparaciones a priori es que, a medida que aumenta el número de contrastes, también se incrementa la probabilidad de cometer un error de tipo I, aunque existen diversos métodos (como la ya citada corrección de Bonferroni o prueba de Dunn) que permiten solventar dicho problema.

El **método de Dunnet** es el método considerado como más adecuado cuando se pretende comparar un grupo de control frente al resto de condiciones experimentales.

Comparaciones no planificadas o a posteriori

Las comparaciones no planificadas, por el contrario, se formulan en función de los resultados obtenidos en el análisis de la varianza y se llevan a cabo para extraer la máxima información posible de los datos del experimento. Estas comparaciones que persiguen el objetivo de mantener constante la probabilidad de cometer un error de tipo I en las decisiones estadísticas, siendo esta su principal ventaja frente a las comparaciones a priori anteriormente citadas.

Encontramos numerosas pruebas entre dichas estrategias:

– El método de las diferencias honestamente significativas de Tukey **(HSD de Tukey**), considerado el más potente cuando el investigador desea realizar todas las comparaciones posibles entre todos los pares de grupos (comparaciones entre pares de medias, siempre y cuando tales comparaciones sean simples.

– El **método de Scheffé**.

– La **prueba Newman-Kleus** es un procedimiento de comparaciones múltiples, muy parecida al método Tukey, pero a diferencia de éste, por el procedimiento que lleva a cabo para sus cálculos, es más probable que revele diferencias significativas entre medias y que comenta más error tipo I **(PIR 20, 8)**. Otros métodos similares al método de Newman-Keuls para las comparaciones múltiples a posteriori cuyo control de alfa no es adecuado son el método de Fisher y el método de Duncan.

– El **método de Peritz**, considerado como el más potente y robusto de todos.

7.4. EFECTOS FACTORIALES

Mediante el Análisis de Varianza el experimentador puede comprobar la influencia de los factores aisladamente, así como la influencia de su interacción. Existe, no obstante, otro procedimiento de contraste-verificación de hipótesis en un diseño factorial: el análisis de los **efectos factoriales**. Por ***efectos factoriales***, entendemos la influencia específica ejercida por los factores sobre las respuestas de los sujetos en la variable dependiente.

Este procedimiento permite inferir la influencia de los factores aislados y sus interacciones en sus diversas modalidades. Es un procedimiento **confirmatorio** del Análisis de Varianza, si la F procedente de un factor o interacción resultó estadísticamente significativo, el análisis de los efectos factoriales tendrá que confirmar y no contradecir este punto. Se trata también de un análisis **complementario**, pues una F significativa pone de relieve que el factor o interacción influye en los datos, pero no especifica qué niveles de la fuente (o combinaciones de niveles en el caso de interacción) producen esa influencia.

Existen dos bloques de efectos factoriales:

– Los que hacen referencia a los efectos producidos por los factores aisladamente considerados que, a su vez, se subdividen en:

  - **Efectos principales**.
  - **Efectos diferenciales**.

– Los que hacen referencia a la acción combinada de dos o más factores, distinguiéndose:

- **Efectos de interacción**.
- **Efectos simples**.

Simplificando mucho, podríamos decir que los efectos principales y los efectos de interacción serían análogos a realizar un ANOVA, mientras que los efectos diferenciales y efectos simples, serían análogos a hacer pruebas a posteriori.

La ventaja más importante que presentan los diseños factoriales, frente a los diseños unifactoriales es que, además de permitir el análisis de los efectos principales, o de la influencia que ejerce sobre la VD cada uno de los factores con independencia del resto, también posibilitan examinar los efectos de interacción entre tales factores. En un diseño factorial siempre será más relevante de cara al análisis los efectos conjuntos de los factores que los efectos por separado. Por ejemplo, cuando la interacción es estadísticamente significativa, los efectos principales no son consistentes y se aconseja interpretar los efectos simples.

RECUERDA

¡Cuidado! En ANOVA debes recordar que, si bien este realiza sus hipótesis sobre la igualdad de **medias**, a nivel estadístico **emplea las varianzas** en sus cálculos.

Efectos principales

El efecto principal de un factor se define como el efecto medio de ese factor bajo todos los niveles del otro. Otra definición que podemos encontrar es la diferencia existente entre la media de un nivel cualquiera de un factor y la media general.

Suele representarse mediante la letra griega, minúscula, correspondiente a la misma letra, en castellano, conque es designado el factor al que hace referencia: (α) efectos principales del factor A; (β) efectos principales del factor B.

Con un ejemplo lo comprenderemos mejor, supongamos un diseño factorial A x B:

	a_1	a_2	
b_1	$\bar{Y}\ a_1 b_1$	$\bar{Y}\ a_2 b_1$	$\bar{Y}\ b_1$
b_2	$\bar{Y}\ a_1 b_2$	$\bar{Y}\ a_2 b_2$	$\bar{Y}\ b_2$
	$\bar{Y}\ a_1$	$\bar{Y}\ a_2$	$\bar{Y}$ ***Gral.***

De acuerdo con la tabla, los efectos principales del factor A se calcularían hallando la diferencia entre la media general y los valores medios de la variable dependiente bajo cada uno de los niveles del factor A:

Efectos principales del factor A (α):

$$\bar{Y}\ a_1 - \bar{Y}$$
$$\bar{Y}\ a_2 - \bar{Y}$$

Las dos diferencias nos expresan respectivamente la influencia de cada uno de los niveles del factor A con relación al resto de los niveles. Si el resultado es cero o la diferencia no es estadísticamente significativa, concluiremos que el nivel concreto del factor no se diferencia de la media del grupo. Hablaremos de efectos principales nulos. Si, por el contrario, la diferencia resultante es significativa, se interpreta que ese nivel concreto ejerce una influencia específica sobre los resultados del experimento.

Como decíamos anteriormente, los efectos principales van a ratificar el significado de la F del Análisis de Varianza, de modo que si la F del factor A resultó significativa, entonces encontraremos que una, al menos, de las diferencias entre los niveles de A y la media general es significativa.

Efectos diferenciales

Los efectos diferenciales de un factor pueden definirse como la diferencia entre los efectos principales de dos niveles de ese factor. Siguiendo con el ejemplo del diseño factorial A x B los efectos diferenciales del factor A se calcularían del siguiente modo:

$$\bar{Y}\,a_1 - \bar{Y}\,a_2$$

Los efectos diferenciales vienen a expresar en qué medida las medias aritméticas obtenidas bajo los distintos niveles de un mismo factor difieren entre sí. Cuanto mayor sea esta diferencia, más clara será la influencia del factor sobre los datos del experimento.

Si la F de un factor es significativa, deberemos hallar, al menos, un efecto diferencial entre dos niveles de ese factor.

Efectos de interacción

Los efectos de interacción responden a la influencia específica de un factor cuando actúa en combinación con otro, es decir, al efecto conjunto de las variables independientes sobre la conducta. Esta interacción puede ser de primer orden, cuando hace referencia a la acción combinada de dos factores, o de segundo orden, si se refiere a la acción de los niveles de tres factores. En el caso de cuatro factores, sería de tercer orden, y así sucesivamente.

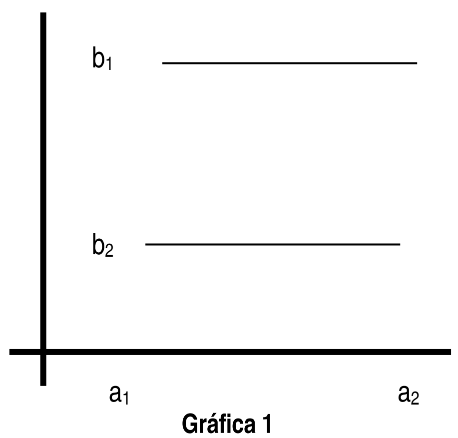
Como señala Kazdin (1992), el efecto de interacción es el resultado de la combinación entre los efectos de varios factores, y su presencia indica que la acción de una variable depende del valor (o valores) que adopta (n) otra (s) variable (s).

El efecto de interacción puede ser estadísticamente significativo sin que lo sean, necesariamente, los efectos principales datos **(PIR 19, 158)**.

Existen dos procedimientos para comprobar la interacción:

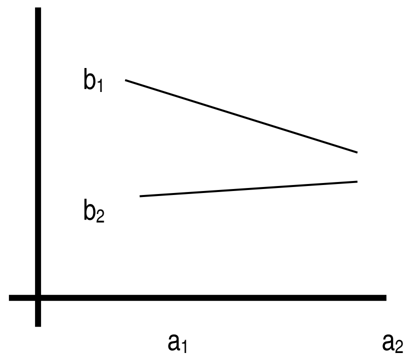
Gráficamente:

En el eje de ordenadas se sitúan las puntuaciones en la variable dependiente (medias), y en el de abscisas, los niveles de uno de los factores, los niveles del factor A, por ejemplo. Se localiza el punto correspondiente a la media aritmética de los niveles de B en su combinación con los de A, y se unen por rectas. Si las rectas son paralelas, indican que no existe interacción entre los factores. Si por el contrario, las rectas no son paralelas, porque se distancian en algún extremo o se cruzan entre sí, ponen de manifiesto que existe interacción entre los factores. Los niveles de un factor, el factor B en nuestro supuesto, modifican su comportamiento, su influencia en los resultados, por el hecho de combinarse con distintos niveles del otro factor.



Gráfica 1

En la gráfica 1 las rectas son paralelas, no existe interacción. Los niveles de B se comportan de la misma manera combinados con el nivel 1 de A como con el nivel 2 del mismo factor.



Gráfica 2

En la gráfica 2 existe interacción entre los factores; las rectas no son paralelas. Significa que los niveles de B modifican su acción por el hecho de combinarse con distintos niveles de A.

Hay ocasiones en que la gráfica no ofrece una información tan clara, siendo imposible discernir si existe o no interacción. En estos casos se recurre al análisis estadístico, que ofrece una cuantía exacta de la interacción y su significación estadística. El procedimiento que más se emplea es el análisis de los efectos simples.

Efectos simples

Los efectos simples vienen a expresar el efecto de un nivel de un factor bajo cada nivel del otro, teniendo por tanto sentido el análisis de estos efectos cuando los efectos de interacción han demostrado ser significativos **(PIR 20, 3)**. En el ejemplo de diseño factorial 2 × 2 que estamos manejando, los efectos simples del nivel a1 consistirán en el efecto de ese nivel bajo cada nivel de B:

$$\bar{Y}\, a_1\, b_1 - \bar{Y}\, a_1\, b_2$$

Y los efectos simples de a_2 serían:

$$\bar{Y}\, a_2\, b_1 - \bar{Y}\, a_2\, b_2$$

Si estas diferencias resultan significativas, concluimos que existe interacción entre un nivel concreto de A y el resto de los niveles del factor B.

Los efectos simples de b_1 y b_2 serían, respectivamente:

$$\bar{Y}\, b_1 a_1 - \bar{Y}\, b_1 a_2$$
$$\bar{Y}\, b_2 a_1 - \bar{Y}\, b_2 a_2$$

RECUERDA

Si los efectos de la interacción son significativos, **no interpretaremos los efectos principales de cada factor, sino los denominados efectos simples.** La interacción puede resultar significativa, aunque los factores por separado no lo sean.

8. ANÁLISIS DE COVARIANZA (ANCOVA): CONCEPTOS BÁSICOS

Sigamos con el ejemplo práctico con el que iniciábamos este tema, pero compliquémoslo un poco más. Supongamos que después de haber construido las tres muestras con pacientes diagnosticados de ataque de pánico y tras haber sometido a cada una de ellas a una técnica terapéutica distinta, descubrimos que el nivel de Neuroticismo medio en la primera muestra es mayor que en la segunda, y a su vez, ésta supera a la tercera. Está claro que las diferencias, si las hay, en el número medio de ataques de pánico por mes van a estar relacionadas con el tipo de técnica aplicada, pero también es posible que estas diferencias estén relacionadas con los distintos niveles de Neuroticismo presentes en las tres muestras. Lo ideal es que esta variable extraña se hubiera tenido en cuenta antes de realizar el experimento, pero incluso ahora no todo está perdido, pues podemos aplicar un control estadístico.

Si la frecuencia de ataques y el nivel de Neuroticismo mantienen una relación lineal, podríamos hacer lo siguiente: restar de los datos que reflejan la frecuencia de ataques aquella parte que es debida a la relación lineal que mantiene con Neuroticismo. Para ello tendríamos que haber medido el nivel de neuroticismo antes de aplicar los tratamientos, para evitar que el tratamiento como tal haya afectado a la variable Neuroticismo. A continuación, tras esta corrección o ajuste, se realizaría un ANOVA.

En resumen, el **ANCOVA** puede ser concebido como un **ANOVA verificado sobre unas puntuaciones corregidas o ajustadas mediante una recta de regresión**. A la variable cuya influencia se desea eliminar la denominamos variable concomitante o covariable.

Además de los supuestos del ANOVA (normalidad, independencia, homocedasticidad), el ANCOVA utiliza unos supuestos propios:

- La covariable no debe estar afectada por el influjo de los tratamientos experimentales (la mejor solución para evitar este problema es tomar la medida antes de aplicar los tratamientos experimentales).

- Los k coeficientes de regresión intragrupo (uno por tratamiento) son homogéneos, es decir, las pendientes de las K rectas de regresión de Y sobre la variable extraña en los K grupos de tratamiento coinciden (supuesto de homogeneidad de las pendientes).

- La regresión de la variable dependiente sobre la covariable es lineal.

El nombre de la técnica se debe a que, al igual que en el ANOVA la variabilidad total quedaba descompuesta en dos componentes aditivos (suma de cuadrados intragrupo y suma de cuadrados intergrupo), en este caso la covariabilidad (o suma de productos) total queda descompuesta o analizada en dos componentes aditivos (suma de productos intragrupo y suma de productos intergrupo).

Sobre ambas sumas de productos se llevan a cabo una serie de transformaciones, con el fin de eliminar de las puntuaciones Y (ataques de pánico) aquella parte que es explicable por el Neuroticismo. El resultado son dos índices:

a) Suma de cuadrados intergrupo corregida.
b) Suma de cuadrados intragrupo corregida.

Una vez que ya hemos corregido el influjo de la variable Neuroticismo, podemos proseguir como si realizáramos un ANOVA, esto es, con las sumas de cuadrados obtenidas derivamos, tras dividirlas por sus grados de libertad correspondientes, sendas medias cuadráticas:

a) Media cuadrática intergrupo (corregida).
b) Media cuadrática intragrupo (corregida).

La media cuadrática intragrupo representa la variabilidad error, aquella variación debida a las diferencias entre individuos en un mismo grupo. La media cuadrática intergrupo, en cambio es la variabilidad entre los distintos grupos atribuible a los diferentes niveles del criterio de clasificación.

Si la proporción de MCinter alcanza un determinado nivel, podremos atribuir las diferencias en las puntuaciones al influjo del criterio de clasificación. El criterio es también aquí un estadístico de contraste:

$$F = \text{MCinter}\,(Y_c) \,/\, \text{MCintra}\,(Y_c)$$

Con (Y_c) se indica que se trata de MC efectuadas sobre puntuaciones Y corregidas (se les ha restado la parte debida a la correlación con Neuroticismo).

Si F es mayor o igual que un determinado valor, se rechaza H_0, esto es, no podemos aceptar que el número medio de ataques de pánico mensuales es igual en todos los tratamientos a nivel poblacional; por lo tanto, concluimos que existen diferencias en el número medio de ataques a nivel poblacional en función de la técnica elegida.

RECUERDA

El ANCOVA es el procedimiento de control **estadístico** por excelencia para controlar el sesgo de selección.

9. EL TAMAÑO DEL EFECTO

El tamaño del efecto mide la magnitud de una diferencia o relación en un estudio, indicando su relevancia práctica más allá de la significación estadística. ¿Por qué es importante? Permite saber si un resultado es realmente relevante o solo significativo por el tamaño de la muestra. Se usa en el cálculo de potencia estadística, determinando cuántos participantes se necesitan en un estudio.

La razón F de un modelo ANOVA de un factor nos informa acerca de si la varianza explicada difiere significativamente de la varianza no explicada por tal modelo, o sea de su significación estadística. Su magnitud no solo depende del efecto que tenga el factor, sino también del tamaño muestral. En consecuencia, una F estadísticamente significativa no necesariamente se asocia con un efecto relevante del factor. Para valorar la significación práctica del resultado de un modelo ANOVA es conveniente utilizar **medidas de tamaño o magnitud del efecto** que nos permiten detectar efectos de tratamiento relevantes.

Se definen como cualquier medida estadística que evidencia el grado con el que un evento dado está presente en una muestra, y nos indica el grado en que la hipótesis nula es falsa. Por ejemplo, supongamos que un profesor elabora un material que, si bien consume muchas horas de estudio, pretende ayudar a sus alumnos a mejorar el rendimiento académico. Lógicamente, el profesor parte de la hipótesis de que las notas son superiores si los alumnos utilizan el material al que nos referimos. Para comprobar su hipótesis, el profesor utiliza dos muestras aleatorias de 900 alumnos cada una, obteniendo una nota media igual a 5,5 en el grupo que ha utilizado el nuevo material (Grupo 1), y una nota media igual a 5 en el grupo que no ha utilizado el nuevo material (Grupo 2). Si, tras el contraste de hipótesis, comprobamos que el nivel crítico p vale < 0,01, concluimos que los resultados son significativos (es decir, la media del Grupo 1 es significativamente a la del Grupo 2).

Ahora bien, el incremento de la nota (es decir, el tamaño del efecto) aparentemente es pequeño, por lo que, teniendo en cuenta que el nuevo material consume muchas horas de estudio, ¿merece la pena emplear dicho material? Si tomamos una decisión teniendo en cuenta únicamente que se han obtenido resultados significativos, concluiríamos que sí que merece la pena, pero si valoramos el tamaño del efecto, probablemente podríamos dudar de si merece la pena utilizarlo. Es por ello por lo que, además de valorar la significación estadística, es relevante cuantificar la magnitud de los efectos observados.

Se han distinguido dos tipos básicos de medidas de magnitud del efecto: los indicadores tipo d, que se basan en la diferencia de medias estandarizada, y los indicadores tipo r, que se basan en la correlación y la proporción de varianza explicada (utilizados comúnmente en el contexto de los modelos ANOVA y el diseño de investigación).

Indicadores tipo d (basados en la diferencia de medias estandarizada):

– d de Cohen: este índice no es más que la estandarización de una diferencia en dos medias (p. ej., de un grupo que ha recibido un tratamiento frente a un grupo que no). Para interpretar el resultado de este índice y, teniendo en cuenta que es una medida estandarizada, Cohen propuso una gradación de la magnitud del efecto en: pequeño: d = 0,2, mediano: d = 0,5 y grande: d = 0,8 o superior.

– Otras alternativas similares: g de Hedges y delta de Glass.

Indicadores tipo r (basados en la correlación o proporción de varianza explicada):

– Correlación: su valor es el indicador más simple, para cuya interpretación Cohen propuso las etiquetas bajo (0.10), medio (0,30) y alto (0,50), aunque existen desacuerdos a este respecto.

– Valor del coeficiente de determinación (R2): de mayor interés en el contexto del diseño de investigación, para cuya interpretación se emplean las etiquetas bajo (0,01), medio/moderado (0,09) y alto (0,25).

– Eta cuadrado (η^2): empleado generalmente para más de dos grupos. En ANOVA sería equivalente al coeficiente de determinación de un modelo de regresión. Sin embargo, se ha demostrado que η^2 es un estimador sesgado de la proporción de varianza explicada (tendencia a la sobreestimación). Entre otras limitaciones, está sesgado por el tamaño muestral. El indicador η^2 se considera por tal razón un indicador dependiente de la muestra (caso muestral).

– Omega cuadrado (ω^2): se trata de un estimador menos sesgado por el tamaño muestral, aunque de uso no muy extendido en psicología.

– Otras alternativas: f de Cohen.

– Existen también otras medidas no paramétricas del tamaño del efecto, como, por ejemplo, la delta de Cliff (para dos o más grupos).

Relación con la Potencia Estadística

Un tamaño del efecto grande facilita detectar diferencias reales con menos participantes. Si el efecto es pequeño, se necesita una muestra mayor para obtener conclusiones fiables. El tamaño del efecto es clave para interpretar estudios: mientras el valor p indica si un resultado es significativo, el tamaño del efecto dice cómo de importante es en la práctica. Por ello, debe reportarse siempre en estudios científicos.

El tamaño del efecto cuantifica la magnitud de una diferencia observada en un estudio. En estadística inferencial, este concepto es fundamental para interpretar la relevancia práctica de un hallazgo, independientemente de su significación estadística.

– *Relación con la potencia estadística:* La potencia de una prueba es la probabilidad de detectar un efecto cuando realmente existe. A mayor tamaño del efecto, mayor potencia tendrá el estudio, ya que las distribuciones de medias de los grupos serán más diferenciadas, reduciendo la superposición.

– *Factores que afectan la potencia:* Tamaño de la muestra (muestras grandes reducen la varianza de la distribución de medias, lo que aumenta la potencia), nivel de significación (un alfa menor reduce la probabilidad de error tipo I, pero también disminuye la potencia) y varianza de la muestra (a menor varianza, mayor precisión en la estimación del efecto).

ESQUEMA DE CONTENIDOS

ESQUEMA RESUMEN SUPUESTOS PRUEBAS PARAMÉTRICAS

SUPUESTO	VIOLACIÓN	PRUEBA ESTADÍSTICA	POSIBLE SOLUCIÓN
NORMALIDAD: las puntuaciones en la VD deben seguir distribución normal	No se ajusta a Distribución Normal *(H0: se ajusta a la normal)*	Pruebas de Bondad de Ajuste: X2 Shapiro-Wilk si $n \leq 50$ Kolmogorov-smirnov si $n > 50$; media y varianza conocidas. Si desconocidas Lilliefors.	Transformación logarítmica de los datos.
HOMOCEDASTICIDAD: igualdad de varianzas de las subpoblaciones	Heterocedasticidad: diferencia de varianzas en las subpoblaciones *(H0: igualdad de varianzas)*	Prueba de Barlett Prueba Levene Prueba de Cochran Prueba Brown-Forsythe Prueba F o Hartley de la prueba *De mayor a menor robustez: Prueba Brown-Forsythe, Prueba de Levene y Prueba F*	T de Welch Corrección grados de libertad prueba F Transformación de las puntuaciones (modelo logaritmo neperiano, método de mínimos cuadrados)
INDEPENDENCIA o “no autocorrelación”: independencia de las puntuaciones entre sí	Autocorrelación: las puntuaciones dependen unas de las otras… *(H0: independencia)*	Test de Durbin-Watson Estadístico de Box- Ljung Rachas	La mejor forma de evitarlo es que sea un M.A.S. Si no es así, habría que transformar el modelo (mínimos cuadrados) + probable situación tipo II

Solo se exigen en diseños de medidas repetidas:

SUPUESTO	VIOLACIÓN	PRUEBA ESTADÍSTICA	POSIBLE SOLUCIÓN
ESFERICIDAD: independencia de los errores, la varianza de las diferencias entre todos los pares posibles de observaciones es la misma	Los errores correlacionan, disminuye la varianza residual, aumenta F artificialmente *(H0: esfericidad)*	W de Mauchly	Corrección de F (ajuste g.l.) MANOVA (Análisis Multivariado de Varianza: considera cada una de las medidas repetidas como una V.D. diferente) *Aumenta probabilidad Error tipo I*
ADITIVIDAD: no interacción sujeto x tratamiento	Interacción sj x trat. (aumenta la varianza error) *(H0: aditividad)*	Prueba de no aditividad de Tukey	Transformación logarítmica de los datos *Aumenta probabilidad Error tipo II*

	PARAMÉTRICAS	NO PARAMÉTRICAS
Bondad de ajuste	--------------------------	χ^2/Kolmogorov-Smirnov
Mediana	--------------------------	Prueba de signos o binomial
Independencia	Contraste de correlación (T)	χ^2
Predicción	Contraste de regresión (T)	--------------------------
Media	Z (varianzas conocidas) T (varianzas desconocidas)	--------------------------
Dos medias independientes	T	Mann-Whitney
Dos medias relacionadas	T	Wilcoxon
Más de dos medias independientes	ANOVA intergrupo	Kruskal-Wallis
Más de dos medias relacionadas	ANOVA intragrupo	Friedman
Más de dos medias (controlando el efecto de alguna variable contaminadora)	ANCOVA inter/intragrupo	---------------------------

Contrastes o comparaciones múltiples

PLANIFICADAS/A PRIORI	NO PLANIFICADAS/A POSTERIORI		
Prueba Dunn- Bonferroni Prueba Scheffé Método de Dunnet	HSD de Tukey Método de Peritz + recomendables	Método de Dunnet Grupo control	Método de Newman- Kleus Método de Fisher Método de Duncan No control adecuado α

Efectos Factoriales

	PRODUCIDOS POR CADA FACTOR AISLADAMENTE	PRODUCIDOS POR LA ACCIÓN COMBINADA DE LOS FACTORES
ANOVA ↓	**Efectos principales**: efecto medio del factor bajo todos los niveles del otro	**Efectos de interacción**: los efectos de un factor cambian dependiendo de los niveles del otro
Si significativo ↓	↓	↓
Necesarias pruebas a posteriori (Scheffé, Bonferroni, Tukey…)	**Efectos diferenciales**: diferencia entre los niveles de un mismo factor	**Efectos simples**: efecto de un nivel de un factor bajo cada nivel del otro