Bloque 3 – Tema 14

ANÁLISIS DE LA VARIANZA. PRUEBA F

El objetivo fundamental de la experimentación es estudiar la posible relación de causalidad existente entre dos o más variables. Este estudio representa la elección de las variables que pueden determinar a otras, pero sobre todo, el **establecer una relación de causa-efecto implica el cumplimiento de tres condiciones**:

a) debe existir una **covariación** (correlación) entre la variable causal (VI) y el efecto (VD)
b) temporalmente, **la causa debe preceder al efecto**
c) **no deben existir variables extrañas**, ajenas al modelo, que puedan explicar el efecto.

En suma, experimentar significa modificar de forma intencionada las condiciones normales de trabajo para localizar mejores maneras de proceder y conseguir, a su vez, un conocimiento más profundo sobre el comportamiento de productos y procesos.

Elementos básicos del diseño de experimentos

Ⓐ *Variabilidad de los resultados*

El objetivo de un experimento es medir el efecto de las variables explicativas (o VI,s) sobre una VD, mientras se controlan otras variables que puedan confundir el resultado.
Supongamos que una variable respuesta (VD) **y**, se va a medir como resultado de dos VI,s $x_1 \ y \ x_2$. Las fluctuaciones que se produzcan en esas variables, y la intervención de otras variables que el experimentador desconoce, harán que el modelo se exprese como: $\boxed{\gamma = f(x_1;\ x_2) + \varepsilon}$ siendo ε el error.
Evidentemente, cuanto menor sea el error experimental, más perfecto será el modelo, en el sentido de poder obtener con él unas estimaciones más precisas.

Ⓑ *Interacción*

Se entiende por interacción el efecto que produce la combinación de dos o más VI,s sobre la VD.

Los diseños factoriales

Vamos a fijar unos conceptos importantes de conocer:

- FACTOR → cada una de las variables que pueden influir en la VD o resultado, y por tanto, se incluirán en el plan de experimentación.
- NIVEL → cada uno de los valores que toma el factor en un experimento.
- RESPUESTA → denominación que toma la VD o resultado.

Si para obtener los resultados solo intervienen unos determinados niveles y siempre los mismos, en cualquiera de las observaciones que se realicen, estaremos ante un **modelo de efectos fijos**.
Si de los diferentes niveles que pueden tomar los factores, se coge una muestra de los mismos, estaremos ante un **modelo de efectos aleatorios**.

Análisis de la varianza factorial

El análisis de la varianza resuelve el problema de comparar varios grupos que tienen sus propias particularidades estimándose, en virtud de los resultados, la diferencia que existe entre dichos grupos y la significación o no de dicha diferencia.
El análisis de la varianza considera los datos de los grupos como un conjunto, y a través de las pruebas estadísticas oportunas, decide si los grupos provienen o no de la misma población (o de poblaciones distintas con la misma varianza), y por consiguiente, si los grupos tienen medias significativamente distintas o no.
Puede parecer extraño que un procedimiento que compara las medias se llame análisis de la varianza. Este nombre deriva del hecho de que para probar la diferencia de medias, estamos comparando realmente (es decir, analizando) las variaciones.
El procedimiento de comparar más de dos medias poblacionales se conoce como análisis de la varianza (ANAVA). Para aplicar la ANAVA deben comprobarse tres supuestos teóricos:

a) **independencia de las observaciones** → los datos de los distintos grupos han sido elegidos al azar de una población normal.
b) **igualdad de varianza** (homocedasticidad) → ser supone que los distintos grupos tienen una varianza común, de modo que no existe diferencia significativa entre las varianzas de los distintos grupos.

c) **los errores** → se distribuyen normalmente.

En las fórmulas que vamos a emplear, el término varianza será equivalente a la suma de los cuadrados dividida por los gl. $\left({}^{SC}/_{gl}\right)$

<u>Análisis de la varianza con un factor. Modelo de efectos fijos</u>

El análisis de varianza de un factor trata de contrastar la Ho de que las poblaciones normales, de las cuales se tomaron las muestras, tienen media común, frente a la H_1 de que al menos dos difieren.

$$\boxed{\begin{matrix} H_0: \mu_1 = \mu_2 = \mu_3 \\ H_1: \mu_1 \neq \mu_2 \neq \mu_3 \end{matrix}}$$

El modelo lineal es: $\boxed{x_{ik} = \mu + a_k + e_{ik}}$

Sea un factor A de tratamientos, con K niveles en dicho factor, en cada uno de los cuales se dispone de $n_k = 1; 2; \ldots\ldots; k$ observaciones independientes, seleccionadas de poblaciones normales con varianza común σ^2. Los símbolos:

- x_{ik} → puntuación (x) del elemento (i) en el grupo (k). Los únicos datos que podemos observar.
- μ → promedio de las K medias de las poblaciones
- a_k → media de la variable para el nivel (k) del factor
- e_{ik} → diferencia entre x_{ik} y μ_k (error del modelo lineal)

<u>Ejecución del modelo:</u>

Sea un diseño que pretende comparar tres métodos de aprendizaje. A cada uno de los métodos se destinan sujetos que obtuvieron unas puntuaciones en una determinada prueba, según la tabla:

Método A	Método B	Método C
X1A	X1B	X1C
X2A	X2B	X2C
X3A	X3B	X3C
X4A	X4B	X4C
X5A	X5B	X5C

En cada método hallaremos:

$\sum X_{ik} \rightarrow X_{iA};\ X_{iB};\ X_{iC}$ (sumatorio de los valores de cada grupo). Sumados todos es $\sum X_T$

$\bar{X}_A = \frac{\sum X_{iA}}{n_A} \qquad \bar{X}_B = \frac{\sum X_{iB}}{n_B} \qquad \bar{X}_C = \frac{\sum X_{iC}}{n_C}$ (medias de cada grupo)

$\bar{X}_T = \frac{\sum\sum X_{ik}}{N} = \frac{\sum X_{iA} + \sum X_{iB} + \sum X_{iC}}{N} \quad siendo\ N = n_A + n_B + n_C$ (medias de todas las puntuaciones del diseño)

<u>Suma de cuadrados</u> (Fórmulas)

Suma de cuadrados total:

$$\boxed{SC_T = \sum\sum (X_{ik} - \bar{X}_T)^2} \qquad \text{ó también} \qquad \boxed{SC_T = \sum X_T^2 - \frac{(\sum X_T)^2}{N}}$$

siendo: SC_T → suma de cuadrados total (con todas las puntuaciones del diseño)

$\sum X_T$ → suma de todas las puntuaciones

$\sum X_T^2$ → suma de los cuadrados de todas las puntuaciones

Suma de cuadrados entre-grupos:

$$\boxed{SC_E = \sum \frac{(\sum X_{ik})^2}{n_k} - \frac{(\sum X_T)^2}{N}} = \left[\frac{(\sum X_{iA})^2}{n_A} + \frac{(\sum X_{iB})^2}{n_B} + \frac{(\sum X_{iC})^2}{n_C}\right] - \frac{(\sum X_T)^2}{N}$$

Suma de cuadrados intra-grupos:

$$\boxed{SC_I = \sum X_T^2 - \sum \frac{(\sum X_{ik})^2}{n_k}} = \sum X_T^2 - \left[\frac{(\sum X_{iA})^2}{n_A} + \frac{(\sum X_{iB})^2}{n_B} + \frac{(\sum X_{iC})^2}{n_C}\right]$$

Como fórmula general de la ANAVA, tenemos que la suma de cuadrados cumple: $\boxed{SC_T = SC_E + SC_I}$

SC_E → Suma de cuadrados entre los grupos. Equivale a la varianza de la media de los grupos, tratados éstos como datos individuales. Son las desviaciones de la media de las puntuaciones en cada grupo respecto a la media total.

SC_I → Suma de cuadrados intra (dentro de) los grupos. Son las desviaciones de las puntuaciones en cada grupo respecto a la media del grupo y refleja la dispersión de los datos dentro de cada grupo.

<u>Hipótesis nula</u>: Ho: $\mu_A = \mu_B = \mu_C$ (no existen diferencias significativas entre las medias de los grupos)

<u>Tabla de la ANAVA</u>:

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	Estadístico F
Intergrupos	SC_E	k-1	$MC_E = \frac{SC_E}{k-1}$	$F = \frac{MC_E}{MC_I}$
Intragrupos	SC_I	N-k	$MC_I = \frac{SC_I}{N-k}$	
Total	SC_T	N-1		

<u>Criterio de decisión</u>:
El estadístico F se distribuye con una F de Snedecor, con (k-1) y (N-k) grados de libertad. Cuando el valor de F sea mayor o igual que el valor de F obtenido en tablas (las tablas son de F de Fisher), rechazaremos la Ho al nivel de significación α. Si $F_{tablas} < F_{exp} \rightarrow rechazo\ Ho$. F es siempre positivo y > 1.

<u>Problemas</u>
Dada la complejidad del método ANAVA y para facilitar la comprensión de sus fórmulas y secuencia de cálculo, se presentan de forma íntegra la realización de dos problemas modelo, según que los grupos tengan igual o distinto número de elementos.
1º) Ejem. 14.2 pág. 326. Ver su enunciado en el texto de la asignatura.

Profesor_A	68	72	77	52	42	60	58
Profesor_B	71	78	56	59	63	71	74
Profesor_C	78	81	82	83	70	65	78

Cálculos previos (con calculadora)

$\sum X_A = 429$ $\quad \sum X_B = 472$ $\quad \sum X_C = 537$

$\sum X_A^2 = 27169$ $\quad \sum X_B^2 = 32228$ $\quad \sum X_C^2 = 41467$

$\bar{X}_A = 61{,}285$ $\quad \bar{X}_B = 64{,}428$ $\quad \bar{X}_C = 76{,}714$

$\sum X_T = 1438$ $\quad \sum X_T^2 = 100864$ $\quad \bar{X}_T = 68{,}476$

Cálculo de los sumatorios de los cuadrados (según fórmulas)

$$SC_T = \sum X_T^2 - \frac{(\sum X_T)^2}{N} = 100864 - \frac{(1438)^2}{21} = 100864 - 98468{,}761 = 2395{,}239$$

$$SC_E = \left[\frac{(\sum X_A)^2}{n_A} + \frac{(\sum X_B)^2}{n_B} + \frac{(\sum X_C)^2}{n_C}\right] - \frac{(\sum X_T)^2}{N} = \left[\frac{(429)^2}{7} + \frac{(472)^2}{7} + \frac{(537)^2}{7}\right] - 98468{,}761 =$$

$$= (26291{,}571 + 31826{,}285 + 41195{,}571) - 98468{,}761 = 844{,}666$$

$$SC_I = \sum X_T^2 - \left[\frac{(\sum X_A)^2}{n_A} + \frac{(\sum X_B)^2}{n_B} + \frac{(\sum X_C)^2}{n_C}\right] = 100864 - 99313{,}427 = 1550{,}573$$

Ahora podemos construir la tabla ANAVA:

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	Estadístico F
Intergrupos	844,666	2	422,333	$\frac{422,333}{86,143} = 4,902$
Intragrupos	1550,573	18	86,143	
Total	2395,239	20		

Criterio de decisión: si F(tablas) < Fexp. → rechazo Ho.
Buscamos en tablas de F de Fisher el valor de F con (3-1) y (21-3) gl., y encontramos que F = 3,55
Como 3,%% < 4,902 → rechazo Ho y si hay diferencias significativas entre los métodos de enseñanza.

Pruebas a posteriori:
Como el estadístico F no discrimina entre las muestras, sino que solo contrasta la Ho acerca de su validez, hemos de utilizar la prueba de las comparaciones múltiples de Scheffe, que tiene por estadístico el siguiente:

$$d = \sqrt{(k-1).F_{critica}\,.MC_I}\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

como en este caso $n_A = n_B = n_C = 7$, tenemos que: $d = \sqrt{2.(3{,}55)(86{,}143)}\ \sqrt{\frac{2}{7}}$ = 13,218
diferencias entre las medias de los grupos: d(AB) = 6,143 ; d(BC) = 9,286 ; d(AC) = 15,429 > 13,218
por tanto encontramos que es significativa la diferencia de medias entre los grupos A y C.

2º) Problema
Un equipo educador preocupado por el comportamiento de los alumnos en clase decide comprobar qué tipo de refuerzo puede favorecer una dinámica más positiva e integrada. Se dividen aleatoriamente 50 estudiantes entre los grupos A (alabanzas), B (premios) y C (castigos), de forma que el A lo integran 20 estudiantes, el B tiene solo 10 y el C tienen otros 20. A través de una escala de interacción verbal, se les calificó después de un periodo de tiempo (2 evaluaciones) en una escala del 1 al 15. Se pretende conocer si las diferencias obtenidas son significativas. Las puntuaciones de los grupos se ofrecen en la tabla adjunta.

(A) Alabanzas				(B) Premios		(C) Castigos			
15	10	18	15	19	12	12	11	9	7
12	11	14	17	11	10	8	8	6	3
13	15	12	14	15	9	7	6	10	9
11	13	15	19	13	12	9	13	11	4
17	12	13	14	9	15	5	9	4	7

Cálculos previos (con calculadora)

$\sum X_A = 280 \qquad \sum X_B = 125 \qquad \sum X_C = 158$
$\sum X_A^2 = 4032 \qquad \sum X_B^2 = 1651 \qquad \sum X_C^2 = 1392$
$\bar{X}_A = 14 \qquad \bar{X}_B = 12{,}5 \qquad \bar{X}_C = 7{,}9$
$n_A = 20 \qquad n_B = 10 \qquad n_C = 20$
$\sum X_T = 563 \qquad \sum X_T^2 = 7075 \qquad \bar{X}_T = 11{,}26 \qquad N = 50$

Cálculo de los sumatorios de los cuadrados (según fórmulas)

$$SC_T = \sum X_T^2 - \frac{(\sum X_T)^2}{N} = 7075 - \frac{(563)^2}{50} = 7075 - 6339{,}38 = 735{,}62$$

$$SC_E = \left[\frac{(\sum X_A)^2}{n_A} + \frac{(\sum X_B)^2}{n_B} + \frac{(\sum X_C)^2}{n_C}\right] - \frac{(\sum X_T)^2}{N} = \left[\frac{(280)^2}{20} + \frac{(125)^2}{10} + \frac{(158)^2}{20}\right] - 6339{,}38 =$$

$$= (3920 + 1562{,}5 + 1248{,}2) - 6339{,}38 = 391{,}32$$

$SC_I = SC_T - SC_E$ = 735,62 – 391,32 = 344,30

Construcción de la tabla de ANAVA:

Fuente de variación	Suma de cuadrados	Grados de libertad	Media de cuadrados	Estadístico F
Intergrupos	391,32	2	195,66	$\frac{195,66}{7,325} = 26,71$
Intragrupos	344,30	47	7,325	
Total	735,62	49		

Criterio de decisión: si F(tablas) < Fexp. $\rightarrow$ rechazo Ho
Buscamos en tablas de Fisher, con (3-1) y (50-3) gl., y encontramos que F = 3,20
Como 3,20 < 26,71 $\rightarrow$ rechazo Ho y hay diferencias significativas entre los tipos de refuerzo.

Efectuamos la prueba de Scheffé: $d = \sqrt{(k-1).F_{critica}\ .MC_I}\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$; como en este caso $n_A \neq n_B \neq n_C$, hemos de efectuar tres comparaciones:

$d(AB) = \sqrt{2.(3{,}20)(7{,}325)}\ \sqrt{\frac{1}{20}+\frac{1}{10}}$ = 2,651 la diferencia de medias entre A y B es 1,5 $\rightarrow$ 1,5 < 2,651, luego la diferencia de medias no es significativa

$d(BC) = \sqrt{2.(3{,}20)(7{,}325)}\ \sqrt{\frac{1}{10}+\frac{1}{20}}$ = 2,651 la diferencia de medias entre B y C es 4,6 $\rightarrow$ 4,6 > 2,651, luego es significativa la diferencia de medias entre los métodos de refuerzo B y C

$d(AC) = \sqrt{2.(3{,}20)(7{,}325)}\ \sqrt{\frac{1}{20}+\frac{1}{20}}$ = 2,164 la diferencia de medias entre A y C es 6,1 $\rightarrow$ 6,1 > 2,164, luego es significativa la diferencia de medias entre los métodos de refuerzo A y C

<u>Pruebas a posteriori (ad-hoc)</u>
Con el análisis de varianza se llega a la conclusión de rechazar o no la Ho de igualdad de medias. En caso de rechazarla, se sabe que por lo menos existe una diferencia significativa entre algún par de medias, pero no se sabe entre qué par o pares se encuentran las diferencias.
El objetivo de los métodos de comparación múltiple es aislar las comparaciones entre las medias que, por ser significativas, influyen en la decisión.
Hay dos métodos de comparación:

- Tukey $\rightarrow$ se emplea cuando los tamaños muestrales son iguales y existe homocedasticidad. Es muy potente. (No la vemos en este curso)
- Scheffé $\rightarrow$ no le influye que los tamaños muestrales sean iguales o diferentes. Es menos potente que el anterior. Su expresión es :

$$\boxed{d = \sqrt{(k-1).F_{critica}\ .MC_I}\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

donde: $d \rightarrow$ diferencia mínima entre dos medias para que sean significativas
n_1 y $n_2 \rightarrow$ tamaños de las muestras
$k \rightarrow$ nº de grupos