Conceptos temas 5 a 7

Diseños de investigación y análisis de datos

Grado en Psicología

Tutora: María B. Font Llompart
marfont@palma.uned.es

Índice

Tema 5. ANOVA. Diseños de más de dos grupos independientes

1 factor con más de dos niveles

- Efecto de 3 dosis diferentes de 1 fármaco
- Efecto de 3 fármacos diferentes

Tema 6. ANOVA. Diseños de más de dos grupos con muestras relacionadas (diseños intra-sujetos, de medidas repetidas o de medidas dependientes)

1 factor con más de dos niveles

Los mismos sujetos se someten a las tres condiciones de la tarea de Stroop

Tema 7. ANOVA. Diseños de más de dos grupos independientes. Análisis de varianza de dos factores.

2 factores con más de dos niveles

Los sujetos realizan tres exámenes (diseños, fisiología y desarrollo) en tres momentos diferentes del día (mañana, tarde, noche)

Etapas de un contraste de hipótesis

- 1. Condiciones de la investigación y supuestos que cumplen los datos observados.
  - paramétrico/no paramétrico…
- 2. Formulación de la hipótesis nula y alternativa.
- 3. Estadístico de contraste.
  - *F*
- 4. Regla de decisión
  - A partir del nivel crítico p/p valor (probabilidad de que siendo cierta H_0 se observen unos resultados como los observados en la muestra o más extremos).
- 5. Conclusión (¿H_0 o H_1?)
  - Comparar el nivel crítico p con el nivel de significación (α) Normalmente rechazamos H_0 si $p<0.05$
- 6. Pruebas *posthoc*
  - ¿Entre que par de grupos hay diferencias?
- 7. Interpretación

1. Condiciones de la investigación

Condiciones para ANOVA:

- Puntuaciones medidas en una escala de intervalo o razón
- Independencia de las observaciones
- Normalidad de las distribuciones en todos los niveles
- Homocedasticidad (varianzas iguales en todos los niveles)

ANOVA

Técnica para comparar medias de <u>más de dos grupos</u>

¿Por qué no utilizamos comparaciones t?

1. El número de comparaciones crece a medida que aumenta el número de grupos/niveles

   Nº de combinaciones: $\frac{\alpha(\alpha-1)}{2}$

2. A mayor número de combinaciones mayor probabilidad de cometer error tipo I (α) y por tanto aumenta la probabilidad de rechazar H_0 siendo cierta (p.157)

ANOVA

¿Por qué es un análisis de medias si se le llama análisis de la varianza?

Estudia qué variabilidad corresponde al factor estudiado y cual a los factores extraños (error experimental)

<u>Modelo estadístico</u>: ANÁLISIS DE UN FACTOR

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

μ = *media*

α_i = *componente común a todos los elementos de un determinado nivel del factor*

E_{ij} = *componente del error experimental (incluye los factores no controlados en el experimento)*

ANOVA

Modelo estadístico: ANÁLISIS DE DOS FACTORES

$$Y_{ij} = \mu_T + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

μ = *media total de la población*

α_i = *promedio del efecto del factor A en el nivel i*

β_j = *promedio del efecto del factor B en el nivel j*

$(\alpha\beta)_{ij}$ = *efecto de la interacción para la combinación de tratamientos i (factor A) y j (factor B)*

ϵ_{ijk} = *error experimental asociado a cada puntuación*

Tipos de diseños

DISEÑOS DE EFECTOS ALEATORIOS O MODELO ALEATORIO:

Los niveles son una muestra de los posibles niveles del factor y las conclusiones son para todos ellos Nos daría lo mismo utilizar esa dosis u otras siempre que se pudiera establecer un sentido ascendente o descendente.

DISEÑO DE EFECTOS FIJOS O MODELOS FIJOS:

Conclusiones restringidas a los niveles elegidos. Pueden existir más niveles pero sólo nos interesan estos.

Tipos de grupos

EQUILIBRADOS:

Todos los grupos tienen el mismo número de sujetos.

NO EQUILIBRADOS:

Los grupos tienen diferente número de sujetos.

ANOVA

Las variables independientes que se estudian se llaman **FACTORES** que se dividen en **NIVELES**.

2. Formulación de H_0 y H_1

En los temas 5, 6 y 7 sólo hay una posibilidad, porque sólo se pueden comprar medias:

DISEÑOS DE UN FACTOR:

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

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$ **al menos para un par de μ_i**

El ANOVA sólo te permite concluir que existen o no diferencias entre grupos pero no entre qué par de grupos

2. Formulación de H_0 y H_1

DISEÑOS DE DOS FACTORES:

Para el factor A:

H_0: $\mu_{A1} = \mu_{A2} = \mu_{A3}$

H_1: $\mu_{A1} \neq \mu_{A2} \neq \mu_{A3}$ al menos para un par de μ_i

Para el factor B:

H_0: $\mu_{B1} = \mu_{B2} = \mu_{B3}$

H_1: $\mu_{B1} \neq \mu_{B2} \neq \mu_{B3}$ al menos para un par de μ_i

Para la interacción:

H_0: No existe interacción

H_1: Existe interacción

3. Estadístico de contraste

El estadístico que utilizamos es la *F* de Fisher

3. Tabla
1 factor (IND)

Dividimos la columna SC por la g.l.

F.V.	S.C.	g.l.	M.C.	F
Inter	SC_{inter}	a−1	$MC_{inter} = \frac{SC_{inter}}{a-1}$	$F = \frac{MC_{inter}}{MC_{intra}}$
Intra	SC_{intra}	N−a	$MC_{intra} = \frac{SC_{intra}}{N-a}$	
Total	SC_{total}	N−1		

g.l. del numerador (a−1) ,
g.l. del denominador (N−a)

F.V.: Fuentes de variación
SC: Suma de cuadrados. Mide la variabilidad dentro de cada nivel (SC_{inter}) o la variabilidad debida al error (SC_{intra}).
MC: Media cuadrática. Varianza insesgada/Cuasi-varianza entre grupos (MC_{inter}) o dentro de los grupos (MC_{intra})
a: número de niveles
N: número total de sujetos

3. Tabla
1 factor (DEP)

Dividimos la columna SC por la g.l.

F.V.	S.C.	g.l.	M.C.	F
Factor (A)	SC_A	a–1	$MC_A = \frac{SC_A}{a-1}$	$F = \frac{MC_A}{MC_{(AxS)}}$
Sujetos (S)	SC_S	s–a	$MC_S = \frac{SC_S}{s-1}$	
Error (AxS)	$SC_{(AxS)}$	(as)–a–s+1	$MC_{Axs} = \frac{SC_{(AxS)}}{(as)-a-s-1}$	
Total	SC_T	(as)–1		

s: número de sujetos en cada nivel

g.l. del numerador (a–1) ,
g.l. del denominador
((as)–a–s+1)

3. Tabla
2 factores

Dividimos la columna SC por la g.l.

F.V.	S.C.	g.l.	M.C.	F
A	SC_A	a–1	$MC_A = \frac{SC_A}{gl_A}$	$F = \frac{MC_A}{MC_{S/AB}}$
B	SC_B	b–1	$MC_B = \frac{SC_B}{gl_B}$	$F = \frac{MC_B}{MC_{S/AB}}$
AxB	SC_{AB}	(a–1)(b–1)	$MC_{AB} = \frac{SC_{AB}}{gl_{AB}}$	$F = \frac{MC_{AB}}{MC_{S/AB}}$
Intra (S/AB)	$SC_{S/AB}$	ab(n–1)	$MC_{S/AB} = \frac{SC_{S/AB}}{gl_{S/AB}}$	
Total	SC_T	abn–1		

g.l. del numerador, g.l. del denominador según el caso

F – Fisher

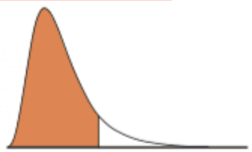
Probabilidad de ser menor que

PROBABILIDAD QUE CAMBIA EN CADA PÁGINA

Grados de libertad n2

Grados de libertad n1

TABLA VII: DISTRIBUCIÓN *F*



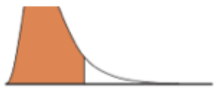
$P(F_{n_1 n_2} \le f_{n_1 n_2}) = 0,90$

Grados de libertad del denominador (n_2)	Grados de libertad del numerador (n_1)															
	1	2	3	4	5	6	7	8	9	10	20	30	40	50	60	120
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195	61,740	62,265	62,529	62,688	62,794	63,061
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,441	9,458	9,466	9,471	9,475	9,483
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,184	5,168	5,160	5,155	5,151	5,143
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,844	3,817	3,804	3,795	3,790	3,775
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,207	3,174	3,157	3,147	3,140	3,123
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937	2,836	2,800	2,781	2,770	2,762	2,742
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,595	2,555	2,535	2,523	2,514	2,493
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,425	2,383	2,361	2,348	2,339	2,316
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,298	2,255	2,232	2,218	2,208	2,184
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,201	2,155	2,132	2,117	2,107	2,082
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,123	2,076	2,052	2,036	2,026	2,000
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,060	2,011	1,986	1,970	1,960	1,932
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,007	1,958	1,931	1,915	1,904	1,876
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	1,962	1,912	1,885	1,869	1,857	1,828
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	1,924	1,873	1,845	1,828	1,817	1,787
16	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055	2,028	1,891	1,839	1,811	1,793	1,782	1,751
17	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028	2,001	1,862	1,809	1,781	1,763	1,751	1,719
18	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005	1,977	1,837	1,783	1,754	1,736	1,723	1,691
19	2,990	2,606	2,397	2,266	2,176	2,109	2,058	2,017	1,984	1,956	1,814	1,759	1,730	1,711	1,699	1,666
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,794	1,738	1,708	1,690	1,677	1,643

Diseños

María Font

F – Fisher



$P(F_{n_1 n_2} \leq f_{n_1 n_2}) = 0{,}90$

ertad del denominador (n_2)	Grados de libertad del numerador (n_1)										
	1	2	3	4	5	6	7	8	9	10	20
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195	61,740
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,441
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,184
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,844
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,207
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937	2,836
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,595
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,425
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,298
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,201
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,123
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,060
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,007
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	1,962

$P(F_{n_1,n_2} \leq f_{n_1,n_2}) = 0{,}95$

ertad del denominador (n_2)	Grados de libertad del numerador (n_1)										
	1	2	3	4	5	6	7	8	9	10	20
1	161,448	199,500	215,707	224,583	230,162	233,986	236,768	238,883	240,543	241,882	248,013
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,446
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,660
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,803
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,558
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	3,874
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,445
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,150
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	2,936
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,774
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,646
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,544
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,459
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,388

Una tabla para cada probabilidad

Diseños
María Font

Propiedad recíproca de la distribución *F*

- Distribución F no es simétrica
- Sirve para calcular probabilidades que no aparecen en la tabla.

$$f_{n1-1,n2-1,\alpha/2} = \frac{1}{f_{n2-1,n1-1,1-\alpha/2}}$$

Aplicación fácil: Calcula el valor correspondiente a una probabilidad de 0,025 con tamaños: $n_1=10$ y $n_2=7$

Aplicación compleja: Pág. 122. Ejercicio 4

4. Regla de decisión

A partir del nivel crítico p/p valor

p-valor: probabilidad de que siendo cierta H_0 se observen unos resultados como los observados en la muestra o más extremos.

IMPORTANTE: Aunque el contraste es bilateral , siempre dejaremos todo el nivel de significación (α) a la derecha. CONSECUENCIA: No hay que multiplicar x2 el valor de la probabilidad para obtener el p-valor.

5. Conclusión

¿H_0 o H_1?

Comparar el nivel crítico p con el nivel de significación (α). Opciones:

- **Si $p>\alpha$** nos quedaremos con H_0. Concluiremos que NO existen diferencias significativas entre los niveles. YA HEMOS TERMINADO.
- **Si $p< \alpha$** rechazaremos H_0 y nos quedaremos con H_1 Concluiremos que existen diferencias significativas entre los niveles. PERO no sabemos entre que par de niveles.

6. Pruebas *post hoc*

- Una vez comprobamos que existen diferencias significativas entre los grupos queremos averiguar entre que par de grupos. Para ello realizamos pruebas post hoc. P.ej. La prueba de las comparaciones múltiples de *Scheffé*

$$CR_{Scheffé} = \sqrt{(a-1) \cdot F_{(1-\alpha),(a-1),(N-a)}} \cdot \sqrt{MC_{Intra}\left[\sum \frac{c_i^2}{n_i}\right]}$$

¿Cómo se calcula?

6. Pruebas *post hoc*

- c_i: Coeficiente de las combinaciones lineales entre distintas medidas a comparar.

	a_1	a_2	a_3	Medias a comparar
Coeficientes	1	-1	0	$\bar{Y}_1 - \bar{Y}_2$
	1	0	-1	$\bar{Y}_1 - \bar{Y}_3$
	0	1	-1	$\bar{Y}_2 - \bar{Y}_3$

Para *Scheffé*

La suma de cada fila de "c" tiene que dar **0**

6. Pruebas *post hoc*

- Si los diseños son equilibrados y se comparan los niveles 2 a 2, es suficiente con calcular una vez el valor del CR de *Scheffé* ya que en todos los casos valdrá lo mismo.
- Una vez calculado comparar con el valor absoluto de la diferencia de medias.
  - Si la diferencia de medias es mayor que el valor de *Scheffé*: Existen diferencias entre esos dos niveles
  - Si la diferencia de medias es menor que el valor de *Scheffé*: No existen diferencias entre esos dos niveles

Para comparaciones dos a dos: Ver ejemplo 5.5 (p.178-180)

Para otras comparaciones: Ver ejemplo 5.6 (p.181-182)