

Selección del tamaño de la muestra en ANOVA un factor

Francesc Carmona
Departament d'Estadística

27 de mayo de 2015

1. Introducción

Un aspecto importante del diseño de un experimento es saber cuántas observaciones se necesitan para que las conclusiones tengan una cierta precisión y la confianza suficiente. Como es lógico esperar, el tamaño de muestra necesario depende de muchas cosas, incluyendo el tipo de experimento previsto, la forma en que se llevará a cabo, los recursos, la sensibilidad deseada y la confianza exigida.

Por sensibilidad se entiende la diferencia en las medias que el experimentador desea detectar, es decir, el experimento debe ser suficientemente sensible para detectar diferencias importantes entre los distintos tratamientos.

En general, el aumento del número de réplicas aumenta la sensibilidad y hace que sea más fácil de detectar pequeñas diferencias entre las medias. Tanto la potencia como el error de estimación son función del tamaño de la muestra N y función de la varianza del error. El objetivo de las técnicas estudiadas es reducir la varianza residual no explicada y con ello mejorar la potencia de los contrastes de hipótesis y reducir el margen de error en la estimación.

2. El contraste de hipótesis

El objetivo es contrastar la hipótesis de que los efectos de los distintos tratamientos son iguales, frente a la alternativa de que algunos no lo son.

La hipótesis nula de que los efectos son todos iguales se expresa como

$$H_0 : \alpha_1 = \dots = \alpha_a = 0$$

ya que imponemos la restricción $\sum_{i=1}^a \alpha_i = 0$ e implica que los α_i son todos iguales a 0.

Como sabemos, al realizar un contraste podemos cometer un error de tipo I con probabilidad

$$\alpha = P[\text{Rechazar } H_0 | H_0 \text{ es cierta}]$$

o un error de tipo II si

$$1 - \beta = P[\text{Aceptar } H_0 | H_0 \text{ es falsa}]$$

La potencia de un contraste¹ se define como

$$\beta = P[\text{Rechazar } H_0 | H_0 \text{ es falsa}] = P[\text{Rechazar } H_0 | H_1 \text{ es cierta}]$$

es, por tanto, la probabilidad complementaria al error del tipo II.

Se trata de construir contrastes que tengan un nivel de significación fijo que controle α y una potencia máxima (es decir un error tipo II pequeño) cosa que cumplen, por ejemplo, los test UMP (uniformemente de máxima potencia).

En este marco queremos calcular la potencia del test F en el caso de efectos fijos:

$$\beta = P[F > F_\alpha(a-1, N-a) | H_1 \text{ es cierta}]$$

Para calcular esta probabilidad, se necesita conocer la distribución de $F = MS_{\text{Trat}}/MS_E$ cuando la hipótesis nula es falsa. Se puede demostrar que en ese caso, el estadístico F se distribuye con una distribución F no centrada con $a-1$ y $N-a$ grados de libertad y un cierto parámetro de centralidad.

¹Algunos autores prefieren llamar β al error de tipo II y $1 - \beta$ a la potencia.

Ejemplo 2.1

Consideremos el caso en el que tenemos cuatro tratamientos en los que utilizamos cuatro drogas diferentes sobre la presión sanguínea de modo que $a = 4$.

Una posibilidad es que dos de las drogas sean efectivas y dos no. Por ejemplo, dos de ellas proporcionan una presión de 110 y dos de ellas de 120. En este caso, los efectos son $(-5, -5, 5, 5)$ de modo que $\sum_{i=1}^4 \alpha_i = 0$ y $\sum_{i=1}^4 \alpha_i^2 = 100$.

Otra posibilidad es que una droga lleve a una presión de 110, dos de ellas a 115 y una a 120. En este caso los efectos son $(-5, 0, 0, 5)$ de modo que también $\sum_{i=1}^4 \alpha_i = 0$ pero $\sum_{i=1}^4 \alpha_i^2 = 50$.

Si observamos ambas situaciones, resulta que aunque no hay diferencias entre los efectos mínimo y máximo, las cantidades $\sum_{i=1}^4 \alpha_i^2$ son muy distintas.

De los dos casos, el segundo es la combinación más desfavorable ya que con ella se obtiene una menor potencia. Sin embargo como en general no sabemos en qué situación estamos, en vez de tratar de adivinar exactamente todos los valores α_i , podemos simplemente especificar $\delta = \max\{\alpha_i\} - \min\{\alpha_i\}$ que es la máxima diferencia entre los efectos.

Volviendo al cálculo en la situación más desfavorable, resulta que $\sum_{i=1}^a \alpha_i^2 = (\delta/2)^2 \times 2 = \delta^2/2$, ya que excepto los valores extremos, el resto de los efectos son cero.

3. Curvas características de operación

Las curvas características de operación representan la probabilidad de error de tipo II $(1 - \beta)$ en función de un parámetro Φ donde

$$\Phi^2 = \frac{\sum_{i=1}^a n_i \alpha_i^2}{a\sigma^2} = \frac{n \sum_{i=1}^a \alpha_i^2}{a\sigma^2}$$

para un modelo balanceado $n_1 = \dots = n_a = n$ de modo que $N = na$.

La cantidad $\lambda = a\Phi^2$ es el parámetro de centralidad de la F y se presentan habitualmente curvas para los niveles de significación $\alpha = 0,05$ y $\alpha = 0,01$.

Así pues, el parámetro Φ depende de

1. Los valores de los efectos $\alpha_1, \dots, \alpha_a$ o bien de los valores medios μ_1, \dots, μ_a para los que se consideran valores distintos:

$$\mu_i = \mu + \alpha_i$$

2. El valor de σ^2 que también es desconocido, aunque se suele usar un valor que se obtiene mediante una muestra piloto.

3. El número de réplicas n para cada tratamiento.

Así, fijados los valores de los efectos y el valor de σ^2 se puede determinar n para que la potencia sea β . Una forma de hacerlo es buscando en las tablas de curvas características de operación como las que hay en el apéndice V del libro de Montgomery[5].

Como a menudo resulta difícil seleccionar los efectos α_i para cada tratamiento que se quieren usar para determinar el tamaño de la muestra, es más sencillo considerar el valor de la máxima diferencia posible entre los efectos δ . De modo que el valor mínimo de Φ^2 es

$$\Phi^2 = \frac{n\delta^2}{2a\sigma^2}$$

Como es el valor mínimo entonces se obtiene el tamaño muestral adecuado para obtener como mínimo la potencia deseada.

Ejemplo 3.1

Como continuación del ejemplo anterior, el procedimiento para elegir la curva característica de operación adecuada es el siguiente:

1. Se calculan los grados de libertad: $v_1 = a - 1 = 4 - 1 = 3$ y se elige el gráfico de curvas correspondiente. En nuestro caso el de la figura 1.
2. Se fija el haz de curvas correspondiente al nivel de significación $\alpha = 0,05$ elegido.

V. Operating Characteristic Curves for the Fixed Effects Model Analysis of Variance (continued)

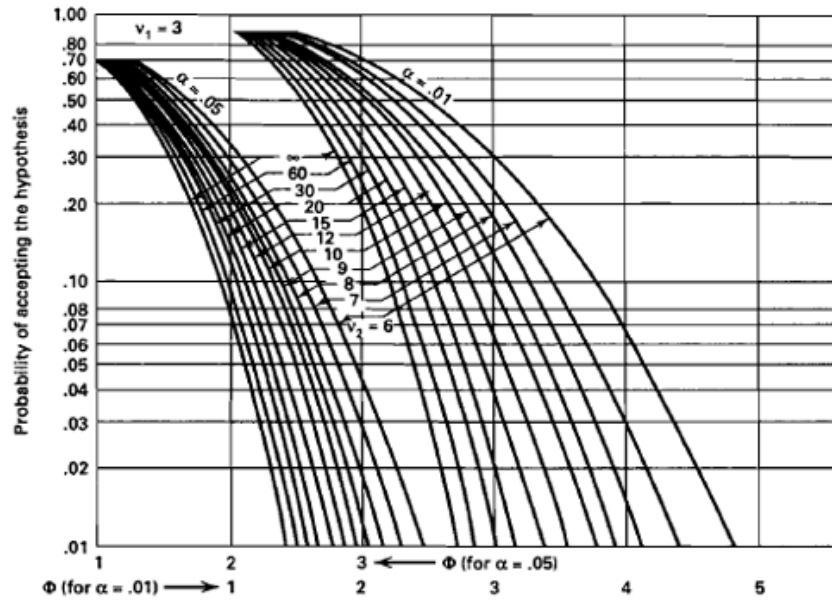


Figura 1: Gráfico de las curvas características de operación para un ANOVA de efectos fijos y 4 tratamientos ($v_1 = 3$).

3. Se elige la curva correspondiente a $v_2 = a(n-1)$. Por ejemplo, si $n = 20$, se tomaría $v_2 = 4 \cdot 19 = 76$.
4. Se calcula el valor del parámetro Φ para unos valores fijos de $\delta = 10$, $\sigma^2 = 144$ (por ejemplo) y el tamaño muestral seleccionado $n = 20$. En este caso $\Phi = 1,3$.
5. Con este valor de $\Phi = 1,3$ en la ordenada y siguiendo la curva seleccionada, se halla una probabilidad de error tipo II de $1 - \beta = 0,45$, luego la potencia es $\beta = 0,55$.

Parece que necesitamos un tamaño muestral mayor.

Bien, vamos a usar un tamaño muestral de $n = 30$. En este caso, tenemos que $\Phi^2 = 2,604$ y $\Phi = 1,6$. Ahora $v_2 = 116$, de modo que el error $1 - \beta$ es 0,30 y la potencia $\beta = 0,70$. Luego necesitamos algo más de $n = 30$ réplicas por grupo para tener un test con una potencia de 0,8.

4. Cálculos con R

4.1. La función `power.anova.test`

Vamos a utilizar la función `power.anova.test` del paquete básico `stats` que permite calcular la potencia de un test ANOVA balanceado de un factor o alguno de los parámetros involucrados a partir de una potencia dada.

En esta función intervienen los siguientes argumentos:

<code>groups</code>	número de grupos o tratamientos
<code>n</code>	número de observaciones (por grupo)
<code>between.var</code>	varianza entre grupos
<code>within.var</code>	varianza dentro de los grupos
<code>sig.level</code>	nivel de significación (probabilidad del error de tipo I)
<code>power</code>	potencia del test (1 menos la probabilidad del error de tipo II)

Exactamente uno de estos parámetros se debe pasar como `NULL` y entonces ese parámetro se calculará en función de los otros.

Ejemplo 4.1

Supongamos que el experimentador está interesado en rechazar la igualdad entre los tratamientos con una probabilidad mínima de 0,9 (error de tipo II: $1 - \beta = 0,1$). Se asumen unas medias poblacionales por grupo igual a

$$\mu_1 = 11, \mu_2 = 12, \mu_3 = 15, \mu_4 = 18, \mu_5 = 19$$

de modo que la media total es $\mu = (11 + 12 + 15 + 18 + 19)/5 = 15$. Supongamos una estimación previa (mediante una muestra piloto) de $\sigma^2 = 9$ y que el nivel de significación α elegido es 0,01. Se tiene que $\alpha_i = \mu_i - \mu$, de manera que

$$\alpha_1 = -4, \alpha_2 = -3, \alpha_3 = 0, \alpha_4 = 3, \alpha_5 = 4$$

Entonces

$$\Phi^2 = \frac{n \sum_{i=1}^5 \alpha_i^2}{5\sigma^2} = \frac{n(16 + 9 + 0 + 9 + 16)}{5 \cdot 9} = 1,11 \cdot n$$

Ahora vamos a calcular la potencia para algunos valores de n :

```
> groupmeans <- c(11, 12, 15, 18, 19)
> power.anova.test(groups = length(groupmeans),
+                   n = 4,
+                   between.var = var(groupmeans),
+                   within.var = 9,
+                   sig.level = 0.01,
+                   power = NULL)
```

Balanced one-way analysis of variance power calculation

```
groups = 5
n = 4
between.var = 12.5
within.var = 9
sig.level = 0.01
power = 0.7065983
```

NOTE: n is number in each group

De modo que la potencia en estas circunstancias es 0,71. Si hacemos los cálculos con otros valores de n obtendremos la siguiente tabla

n	Φ^2	Φ	g.l. ($a - 1, a(n - 1)$)	Potencia	Error tipo II
4	4,44	2,11	(4,15)	0,71	0,29
5	5,55	2,36	(4,20)	0,88	0,12
6	6,66	2,58	(4,25)	0,96	0,04

Por tanto, es necesario realizar un mínimo de 6 réplicas por tratamiento.

4.2. La función `pwr.anova.test`

Otra posibilidad es utilizar la función `pwr.anova.test` del paquete `pwr` que, además de ésta, contiene varias funciones para calcular la potencia de diversos test².

Esta función es muy parecida a la anterior, pero utiliza como parámetros los siguientes:

<code>k</code>	número de grupos o tratamientos
<code>n</code>	número de observaciones (por grupo)
<code>f</code>	tamaño del efecto
<code>sig.level</code>	nivel de significación (probabilidad del error de tipo I)
<code>power</code>	potencia del test (1 menos la probabilidad del error de tipo II)

²Ver Quick-R

donde f es el tamaño del efecto³ de Cohen[2]:

$$f = \sqrt{\frac{\sum_{i=1}^a \frac{n_i}{N} (\mu_i - \mu)^2}{\sigma^2}} = \sqrt{\frac{\sum_{i=1}^a \alpha_i^2}{a\sigma^2}}$$

en el caso balanceado.

Si pensamos en el caso más desfavorable, resulta

$$f = \sqrt{\frac{\delta^2}{2a\sigma^2}}$$

de modo que $\Phi^2 = nf^2$.

Ejemplo 4.2 Como continuación de los ejemplos 2.1 y 3.1, vamos a calcular el tamaño de la muestra necesario para obtener una potencia de 0,8:

```
> pwr.anova.test(k=4, f=sqrt(10^2/2/4/144), sig.level=0.05, power=0.8)
```

Balanced one-way analysis of variance power calculation

```
      k = 4
      n = 32.39305
      f = 0.2946278
sig.level = 0.05
power = 0.8
```

NOTE: n is number in each group

de modo que el número de réplicas por tratamiento debe ser al menos 33.

El valor del tamaño del efecto es cercano a lo que Cohen consideró un valor medio (sobre 0,25).

4.3. La función Fpower1

El paquete `daewr` asociado al libro de Lawson[4] contiene la función `Fpower1` que resulta muy práctica ya que permite calcular la potencia de un rango de valores para el tamaño de la muestra. Esta función utiliza prácticamente los mismos parámetros que las anteriores.

El mismo paquete dispone también de una función `Fpower2` para el caso de dos factores con interacción.

Ejemplo 4.3 Como continuación de los ejemplos 2.1 y 3.1, vamos a calcular la potencia para un rango dado del tamaño de la muestra:

```
> library(daewr)
> Fpower1(alpha=0.05, nlev=4, nreps=30:35, Delta=10, sigma=sqrt(144))
```

	alpha	nlev	nreps	Delta	sigma	power
[1,]	0.05	4	30	10	12	0.7643438
[2,]	0.05	4	31	10	12	0.7798134
[3,]	0.05	4	32	10	12	0.7944621
[4,]	0.05	4	33	10	12	0.8083129
[5,]	0.05	4	34	10	12	0.8213908
[6,]	0.05	4	35	10	12	0.8337222

de modo que el número de réplicas por tratamiento es 33 o más.

³No se debe confundir con el tamaño del efecto η^2 que se utiliza en ANOVA.

Referencias

- [1] F. Carmona, *Modelos lineales*, Publicacions UB, 2005.
- [2] J. Cohen, *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale,NJ: Lawrence Erlbaum, 1988.
- [3] J.J. Faraway, *Linear Models with R*, Chapman & Hall/CRC, 2004.
- [4] J. Lawson, *Design and Analysis of Experiments with R*, Chapman & Hall/CRC, 2015.
- [5] D.C Montgomery *Design and Analysis of Experiments*, John Wiley & Sons, 2008.