Psicothema
2000. Vol. 12, Supl. nº 2, pp. 253-256
ISSN 0214 - 9915 CODEN PSOTEG

Prueba de aleatorización vs. distribución F cuando la escala de medida de la variable dependiente es discreta y el diseño experimental

José Fernando García Pérez, María Dolores Frías Navarro y Juan Pascual Llobell
Universidad de Valencia

El objetivo de este trabajo es analizar cómo incide el número de observaciones en la validez de la conclusión estadística de los diseños experimentales —asumiendo la manipulación y asignación aleatoria— cuando la variable dependiente se mide en una escala ordinal y discreta con una distribución rectangular —los sucesos son equiprobables y finitos. Las variables dependientes que se emplean en la investigación psicológica en bastantes ocasiones no se miden con una variable continua y, en otros casos, aunque la variable sea continua, tampoco es posible asumir que sigan una distribución normal. En este trabajo se construye una distribución muestral del estadístico *F* para una escala ordinal discreta, comparándola con las clásicas que se calcularon asumiendo que la escala de medida de la variable dependiente es continua y normal.

Randomness tests versus F-distribution when the measurement scale is discrete. The aim of this research is to examine the influence of sample size on the statistical validity of experimental designs (i.e. assuming random assignment and direct manipulation) when the dependent variable is discrete and measured in an ordinal scale, which is distributed as a rectangular distribution —any point of the sample space is equally likely. The dependent variables in the psychological studies always are not continuous and, even in this case, frequently, it is difficult to guaranteed the normality assumption. In this paper we constructed a sample distribution of the *F* statistic for the ordinal and discrete scale, and we compare this distribution which the classical distributions which assume that the dependent variable is normally distributed.

Una serie de trabajos cuestionan la validez estadística de las pruebas paramétricas (*F* y *t*) en campos de investigación que habitualmente emplean intensivamente la metodología experimental para contrastar sus teorías. Las dudas aparecen porque la forma de esta distribución no se ajusta a la curva normal. En concreto, *el tiempo de reacción* en las tareas de memoria, atención, decisión léxica, etc. sigue una distribución asimétrica positiva, que puede estar asociada al efecto de algunas variables extrañas como la distracción, o producirse por la desmotivación del sujeto después de un fracaso en la tarea experimental. Estas circunstancias implican que algunas observaciones sean muy extremas (tiempos de reacción muy altos) y la forma final de la distribución muestral resulte, por lo tanto, asimétrica positiva (véase la exposición de Luce, 1986).

Estas puntuaciones extremas aumentan la desviación típica de la distribución de la variable dependiente y, consecuentemente, el error típico de la estimación. Si la prueba de la hipótesis se realiza incluyendo estas observaciones, frecuentemente, la pérdida en la potencia obliga a incrementar considerablemente su número para controlar el error de Tipo II. Por otra parte, es previsible el aumento del error de Tipo I, puesto que la prueba es robusta al incumplimiento de la normalidad, aún en casos mucho más extremos (Por ejemplo: Bock, 1975, p. 111; Borges, Sánchez y Cañadas, 1996; Myers, 1966/1979, p. 72).

Entre las alternativas propuestas y seguidas destacan la eliminación de los datos extremos mediante criterios estadísticos o teóricos (Ratclifs, 1993; Ulrich y Miller, 1994), realizar la transformación logarítmica o inversa de la variable dependiente y aplicar las pruebas de aleatorización (Blair y Karniski, 1993). En general los recortes de los datos, sobre todo, si exceden al 5% de la distribución (Ulrich y Miller, 1994) afectarán a la precisión de la prueba de la hipótesis, de la misma manera que si se emplea la mediana u otros estadísticos diferentes de la media (Miller 1988, 1991). Las transformaciones de la variable dependiente implican necesariamente que la interpretación de los datos con estas nuevas escalas hagan que pierdan los valores y estadísticos el significado original de los tiempos de reacción (Stenberg, 1969). Asimismo, las pruebas de aleatorización conllevan realizar un importante número de cálculos para cada experimento, y, por otra parte, tampoco se ha comprobado que estas pruebas estuvieran, finalmente, libres de cumplir con el supuesto de la independencia (véase Hayes, 1996).

En este trabajo se propone un método alternativo para analizar los resultados de un experimento cuando la variable dependiente es asimétrica y existen puntuaciones extremas. El proceso consiste en cambiar la forma de la distribución de la variable dependiente por otra rectangular donde cada valor en la nueva escala sea equiprobable. Aplicando el análisis de la varianza con los valores de la variable dependiente de la nueva escala.

Correspondencia: José Fernando García Pérez
Facultad de Psicología
Universitat de València
46010 Valencia (Spain)
E-mail: Fernando.García@uv.es

El cambio de la escala, para este caso, consistiría en determinar un número de cortes equivalentes en la distribución original; en el ejemplo que analizamos habría que realizar cinco (a partir de los percentiles: 16.67, 33.33, 50.00, 66.67, 83.33) para obtener 6 intervalos equivalentes. Para determinar el número de intervalos habría que determinar el error típico de estimación, garantizando, como mínimo que las bandas establecidas no se sobrepongan (Martínez, 1995, p. 627). En la nueva escala los intervalos fijados mantienen la relación de orden con la original, pero no se puede establecer la dependencia lineal entre las dos.

Para realizar la prueba de la hipótesis es preciso conocer la función de distribución muestral del estadístico; como ésta variará con el número de intervalos que se establezcan, es necesario calcularla considerando los parámetros: tamaño de la muestra y número de intervalos. Es un caso análogo a una prueba aleatorización (no se emplea el término permutación porque, como luego se verá, se supone la aleatorización). No obstante cabe suponer que a efectos prácticos, por lo que enuncia el teorema del límite central, cuando la muestra sea grande la distribución muestral sea aproximada a la de las clásicas tablas de la *F* (Snedecor, 1934), calculadas asumiendo que la población padre de la que se realiza la extracción es normal (Kirk, 1995, p. 51).

El Teorema del límite central: la media de las muestras

Para ilustrar el efecto que tiene el teorema del límite central en la función de probabilidad de un estadístico trataremos el tema de la media muestral en el caso de que se extraigan dos muestras de una población padre de seis sucesos equiprobables (rectangular). La población muestral, todas las posibles muestras que pueden obtenerse en este caso, son las que se presentan en la Tabla 1.

Este supuesto de independencia implica que la probabilidad de obtener cualquiera de los elementos de la población es independiente y constante, no existiendo ninguna probabilidad condicional entre ellos. Para cumplir este supuesto es preciso que se produzca la asignación aleatoria, asimismo, para que el efecto de la extracción no altere la función de probabilidad de la distribución padre, también es necesario que el número de elementos de la misma sea infinito. El tiempo de reacción que se mida en un experimento puede asumirse que tiene un número ilimitado de elementos puesto que se trata, obviamente, de un proceso que no se agota (Pedhazur y Schmelkin, 1991: 322 y ss.; Fritz, 1998).

Mientras que la función de distribución padre es rectangular, la muestral tiene una forma que se aproxima a la de la normal (Véase la Figura 1). El grado de aproximación aumentará, conforme se aumenten el número de observaciones de cada muestra.

Asimismo, la media de las medias muestrales es la misma que la de la población padre, y la varianza de la *distribución muestral* de la media (σ^2_{media}) *n* veces menor que la varianza de la *distribución padre* (σ^2). Siendo *n* el tamaño de las muestras de dicha *distribución muestral* (Padilla, Merino, Rodríguez-Miñón yMoreno, 1996: 394-399).

Función de distribución del estadístico *F* con 6 intervalos equiprobables en la variable dependiente

Para calcular la función de probabilidad del estadístico *F* supondremos todos los resultados posibles que pueden producirse cuando se realiza el análisis de la varianza con dos muestras de dos elementos cada una, extraídas de esta población padre. En primer lugar hay que calcular la función de distribución muestral de las

Tabla 1
Población muestral de la media de dos muestras extraídas aleatoriamente de una distribución padre rectangular de 6 sucesos

	1	**2**	**3**	**4**	**5**	**6**
1	1.0	1.5	2.0	2.5	3.0	3.5
2	1.5	2.0	2.5	3.0	3.5	4.0
3	2.0	2.5	3.0	3.5	4.0	4.5
4	2.5	3.0	3.5	4.0	4.5	5.0
5	3.0	3.5	4.0	4.5	5.0	5.5
6	3.5	4.0	4.5	5.0	5.5	6.0

Tabla 2
Función de distribución muestral de las sumas de cuadrados del error y del factor A

SC_{Error}^{i}	n^{i}	$SC_{Error}^{i}\,n^{i}$	SC_{A}^{i}	n^{i}	$SC_{A}^{i}\,n^{i}$
0.00	36	0	0.00	146	0
0.50	120	60	0.25	280	70
1.00	100	100	1.00	250	250
2.00	96	192	2.25	208	468
2.50	160	400	4.00	160	640
4.00	64	256	6.25	112	700
4.50	72	324	9.00	70	630
5.00	120	600	12.25	40	490
6.50	96	624	16.00	20	320
8.00	48	384	20.25	8	162
8.50	80	680	25.00	2	50
9.00	36	324	Σ	1296	3780
10.00	64	640			
12.50	72	900			
13.00	40	520			
14.50	32	464			
16.00	16	256			
17.00	24	408			
20.50	16	328			
25.00	4	100			
Σ	1296	7560			

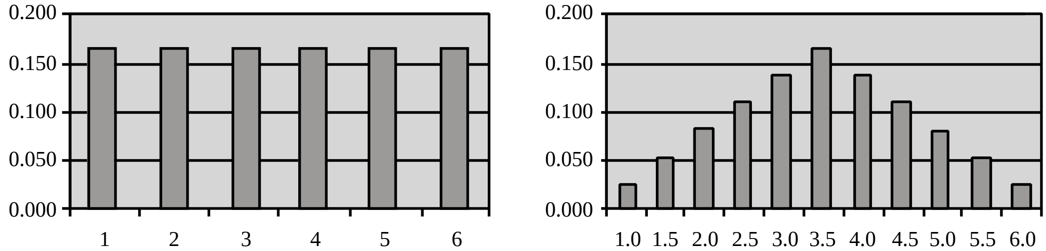

Figura 1. *Función de probabilidad de la distribución padre y de la media de dos observaciones extraídas aleatoriamente*

sumas de cuadrados del término residual y del error (Tabla 2). El número completo de muestras independientes de 2 elementos que pueden resultar en cada una de las dos extracciones será justo el número de variaciones con repetición de 6 elementos tomados de dos en dos ($6^2 = 36$). Como se extraen dos muestras de dos elementos cada una, simulando el análisis de un experimento en que hubiese un factor con dos condiciones experimentales, podemos saber que la primera muestra será una de las 36 de la Tabla 1, y que la segunda, otra de éstas —incluyendo entre las posibles, claro está, a la misma que se obtuvo en la primera extracción—, existiendo en total 1296 resultados posibles ($36^2 = 1296$). Como es factible conocer todos los resultados que pueden obtenerse, se puede saber, en el caso de que no exista un efecto real debido al tratamiento, en cuantas ocasiones se concluiría erróneamente que sí lo hay; esto es, se cometería un error de *Tipo I*.

En un modelo elemental con un *factor A* de *a* condiciones y las mismas *n* observaciones por condición de *a*. Si las $n \cdot a$ observaciones son extraídas independientemente de una misma población de media μ y varianza σ^2 se sabe que (Hays, 1981: 625 y ss.):

$$E[MC_{\mathrm{A}}] = E[MC_{\mathrm{Error}}] = \sigma^2$$

Como las dos distribuciones corresponden a la población muestral de las *Sumas de Cuadrados* podemos calcular directamente el valor esperado o promedio de cada población conociendo las *m* sumas de cuadrados y la frecuencia (n_{i}) de cada i*ésimo* elemento:

$$E\left[SC_{Error}\right] = \frac{\sum_{i=1}^{a} SC_{Error\ j}\, n_j}{\sum_{i=1}^{a} n_j} = \frac{7560}{1296} = 5.833333$$

$$E\left[SC_{A}\right] = \frac{\sum_{i=1}^{a} SC_{Aj}\, n_j}{\sum_{i=1}^{a} n_j} = \frac{3780}{1296} = 2.916667$$

Como en el término de error hay dos grados de libertad ($gl_{Error} = N - \mathrm{a} = 4 - 2 = 2$) y en el factor *A* uno ($gl_{\mathrm{A}} = \mathrm{a} - 1 = 2 - 1 = 1$), podemos comprobar como las dos medias cuadráticas son estimadores insesgados de la varianza de la *población padre*:

$$E\left[MC_{Error}\right] = E\left[\frac{SC_{Error}}{gl_{Error}}\right] = \frac{1}{gl_{Error}} E\left[SC_{Error}\right] = \tfrac{1}{2} \cdot 5.833333 = 2.916667$$

$$E\left[MC_{A}\right] = E\left[\frac{SC_{A}}{gl_{A}}\right] = \frac{1}{gl_{A}} E\left[SC_{A}\right] = 1 \cdot 2.016667 = 2.916667$$

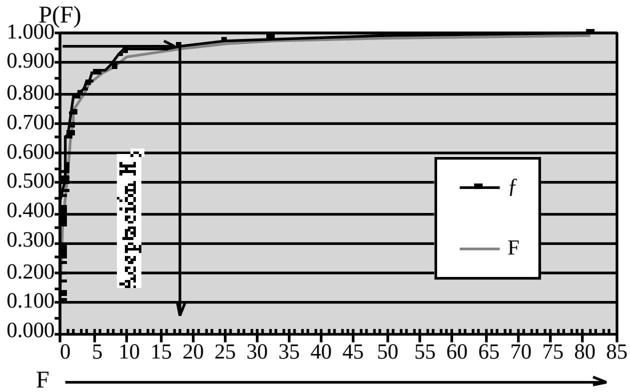

Figura 2. *Función de distribución muestral de $f_{1,\ 2}$ a partir de la población padre discreta y $F_{1,\ 2}$ asumiendo que la distribución padre es normal*

Cumpliéndose que:

$$E[MC_{\mathrm{A}}] = E[MC_{Error}] = \sigma^2 = 2.916667$$

A continuación se calculan todos los posibles valores de *F* —emplearemos la tipografía *f* para distinguirla del estadístico que hemos calculado asumiendo que la *distribución padre* es normal, *F*— que pueden obtenerse cuando se haga la prueba de la hipótesis y se acumulan para construir la función de distribución muestral con un grado de libertad en el numerador y dos en el denominador, cuando se extraigan los elementos de esta distribución padre (Figura 2).

Como puede apreciarse en la Figura 2 y en la primera fila de la Tabla 3, el 95% de los valores de *f* están por debajo de 18.00. Redondeando los cálculos a dos decimales, el valor de α coincide con el de la distribución muestral de *F* cuando la *población padre* es *normal*. Aunque esta coincidencia encontrada en el percentil 95 no implica, como se aprecia en la gráfica, que las dos distribuciones sean iguales. Comprobando todos los puntos estimables de la *distribución padre* discreta con los de la normal, el promedio de las discrepancias en las ordenadas de las dos distribuciones en términos absolutos —lo hemos denominado en la Tabla 3 $|f - \mathrm{F}|$— es de un 3%.

Realizando el mismo proceso aumentando el número de extracciones de cada muestra, las diferencias entre las dos distribuciones disminuye; de tal manera que con dos grupos de 5 observaciones, está en 0.006 (Véase Tabla 3). Como ocurre con estos procesos de cálculo, cuando el experimento se realiza con 5 grados de libertad en el denominador el número de combinaciones ya es de más de 60 millones, y más allá de este número los cálculos y recuentos resultan prácticamente inviables a no ser que se recurra al cálculo diferencial. Pero la diferencia encontrada entre las distribuciones muestrales es pequeña, siendo posible preservar la validez estadística del cálculo dentro de unos límites razonables sin necesidad de tener que realizar los cálculos.

Conclusiones

Parece ser que la validez estadística del procedimiento propuesto es aceptable, puesto que permite estimar los valores exactos de las distribuciones muestrales de los estadísticos. No obstante, si los grados de libertad del término del error son muy pocos (a partir de 6 grados la diferencia para la distribución estudiada es inferior a una centésima), se puede emplear las tablas clásicas de la distribución *F* sin que la pérdida de precisión sea muy considerable. A la hora de determinar una cifra puede tomarse un número entre 10 y 20, tal y como recomienda Bock (1975: 111).

Sí que es necesario poder suponer el principio de aleatorización para garantizar la validez de los resultados, aunque la mayoría de los experimentos que emplean como variable depen-

Tabla 3

Discrepancias entre las distribuciones muestrales del estadístico *f* de una población padre discreta y normal, F

n_i	gl_A	gl_{Error}	$6^n \cdot 6^n - 36$	f	$1 - \alpha_f$	$1 - \alpha_F$	$\lvert f - F \rvert$
2	1	2	1,260	18.000	0.95	0.95	0.028
3	1	4	46,620	9.000	0.95	0.96	0.012
4	1	6	1, 679, 580	6.231	0.95	0.95	0.009
5	1	8	60,466,140	5.565	0.95	0.95	0.006

diente el tiempo de reacción son experimentales en el sentido que cumplen con la manipulación, respecto de la asignación aleatoria en la mayoría de los casos es cuestionable porque se trata de diseños de medidas repetidas. Para garantizar la validez de conclusión estadística es necesario que se corrija la distribución muestral del estadístico de acuerdo con el grado que se desvíen del supuesto de esfericidad o que apliquen la solución multivariada que existe para este diseño.

Referencias

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*, 518-524.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Borges, A., Sánchez, A., & Cañadas, I. (1996). El contraste de medias con grupos pequeños, con escalas ordinales y en ausencia de normalidad. *Psicológica*, *17*, 455-466.

Fritz,R. W. (1998). Interpreting statistical testing: process and propensity, not population and random sampling. *Behavior Research, Methods, Instruments, and Computers*, *30*, 527-535.

Hayes, A. F. (1996). Permutation test is not distribution-free: testing Ho: p = 0. *Psychological Methods*, *1*, 184-189.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Kirk,R.E.(1995). *Experimental design: procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole Publishing.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

Martínez, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis Psicología.

Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Perfomance*, *14*, 539-543.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *Quaterly Journal of Experimental Psychology*, *43A*, 907-912.

Myers, J. L. (1966/1979). *Fundamentals of experimental design*. Boston: Allyn and Bacon.

Padilla,M.,Merino, J. M.,Rodríguez-Miñón, P., & Moreno,E.(1996). *Psicología matemática I*. Madrid: UNED.

Pedhazur, E. J., & Schmelkin, L. (1991). *Measurement, design and analysis*. Hillsdale, NJ: LEA.

Ratclifs, R. (1993). Methods for dealing with reaction times outliers. *Psychological Bulletin*, *114*, 510-532.

Snedecor, G. W. (1934). *Calculation and interpretation of analysis of variance and covariance*. Monograph n.1, Iowa State College, Division of Industrial Science. Ames, IW: Collegiate Press.

Stenberg, S. (1969). The discovery of processing slags: Extensions of Donder's method. En W. G. Koster, ed., *Attention and performance, 11*, (pp. 267-315). Amsterdang: North-Holland.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*, 34-80.