DETERMINACION POR COMPUTADORA DEL TAMAÑO DE MUESTRA Y POTENCIA ESTADISTICA EN ENSAYOS AGROPECUARIOS

MARIA. V. LOPEZ[1], A. GARSD[1], A. SEPLIARSKY[1], H. GIORGINI[1], F. CABONA[1]

Recibido: 01/12/93
Aceptado: 08/07/94

RESUMEN

El presente trabajo considera la potencia estadística en la planificación de los ensayos agronómicos. A partir de una ecuación básica se calculan el tamaño de muestra, la potencia y la diferencia entre parámetros referidos a hipótesis que involucran promedios o proporciones, con una o dos muestras, independientes o apareadas, con aplicación a ejemplos reales. Se incluye, además, el diagrama de flujo del programa de computación adaptable a los distintos tipos de ejemplos.

Palabras clave: tamaño de muestra, potencia estadística, ensayo agrícola, prueba de hipótesis.

COMPUTER DETERMINED SAMPLE SIZE AND STATISTICAL POWER IN AGRICULTURAL EXPERIMENTS

SUMMARY

This paper considers the problem of the statistical power required for planning agricultural and animal production experiments. A general method is presented from which specific equations are derived in the context of real examples. Sample size and power are calculated for the significance of independent and paired differences between parameters from hypotheses involving means and proportions. The article also includes a flowchart of a computer program adaptable to the examples considered.

Key words: sample size, statistical power, agricultural and animal production experiments, hypothesis testing.

INTRODUCCION

Una de las tareas más frecuentes a la que se enfrentan los investigadores en los ensayos agronómicos es la de probar la significatividad de una diferencia entre dos promedios y así poder detectar el verdadero efecto de una nueva tecnología aplicable a la producción, para mejorarla y aumentar su rentabilidad.

Muchos ensayos agronómicos involucran un gran esfuerzo en tiempo y dinero, y sin embargo no llegan al objetivo propuesto. Ocasionalmente se observan diseños inapropiados desde su inicio, sencillamente porque el investigador no ha considerado la probabilidad de fracasar en el intento de descubrir efectos

[1]Cátedra de estadística. Facultad de Agronomía. Universidad de Buenos Aires, San Martín 4453, (1417) Buenos Aires, Argentina.

verdaderos. En algunos casos, los resultados apuntan en la dirección correcta, pero carecen de significatividad estadística. De tal modo, al no encontrarse diferencias significativas existen dos posibilidades: o realmente no hay diferencias o, si las hay, el diseño del ensayo no ha sido efectivo para evidenciarlas.

Una de las formas de controlar la ambigüedad inherente a un ensayo inconclusivo es a través de la consideración del tamaño de la muestra empleada, cuyo cálculo permite al investigador prevenirse del error de no dar cuenta de una diferencia real. La consideración de la probabilidad de no cometer este tipo de error -la potencia en la terminología estadística- se convierte en una precaución fundamental en el momento del diseño.

El objetivo del presente trabajo es formular una expresión sencilla y flexible para el cálculo del tamaño de muestra y de la potencia con el empleo de un programa con aptitud para adaptarse a diferentes tipos de situaciones. El trabajo se ciñe a ensayos comparativos de hasta dos tratamientos.

En la sección 2 se presentan conceptos estadísticos básicos referentes a la prueba de hipótesis describiendo, además, un programa simple y al alcance de cualquier investigador que permite realizar los cálculos planteados. En la sección 3 se consideran los casos particulares de comparaciones de promedios y proporciones para muestras independientes y relacionadas. Se incluyen en esta sección ejemplos agronómicos reales. Finalmente, en la sección 4 se discuten las expresiones presentadas.

El cálculo del tamaño de muestra para ensayos biológicos comparativos de uno y dos tratamientos con consideración de la potencia estadística ha sido tratado ampliamente en la bibliografía. Cohen (1977), Snedecor y Cochran (1979), Steel y Torrie (1980) y Lachin (1981) son algunos ejemplos.

FUNDAMENTOS ESTADISTICOS

Todo ensayo comparativo requiere de la formulación de hipótesis que deben ser probadas a partir de los datos de la muestra.

La estructura estadística de una prueba de hipótesis comienza con el planteo de dos conjuntos posibles de valores del parámetro de interés en correspondencia con dos hipótesis: la hipótesis nula (H_0), típicamente especificada en forma exacta y expresando la ausencia del efecto del tratamiento que se investiga, y la hipótesis alternativa (H_1), especificada como un conjunto de valores del parámetro que prevalecería si el tratamiento ejerciera algún efecto.

Supóngase, por ejemplo, que la proporción P de preñez de vaquillonas al primer servicio en cierta región es de 0,70 y que se desea probar si aplicando una nueva técnica de manejo de los animales se lograría mejorar ese porcentaje. Las hipótesis en este ntes:

$$H_0: P \leqslant 0{,}70$$
$$H_1: P > 0{,}70$$

H_0

$H_1: P > 0{,}70$

Para probar la hipótesis nula de este ejemplo se utiliza la proporción muestral p. Este estadístico tiene una distribución aproximadamente normal cuando el tamaño de la muestra N es suficientemente grande (Chou, 1980). Los parámetros de la distribución son la proporción media P_0 y la variancia σ_0^2/N, donde el subíndice o indica el valor especificado en H_0. La distribución de p de acuerdo con la hipótesis alternativa depende de la proporción poblacional P_1 de vaquillonas preñadas que se espera lograr con el nuevo manejo y la variancia $\sigma_1{}^2/N$, donde el subíndice 1 indica el valor especificado en H_1.

La ocurrencia de un valor de p poco probable bajo la distribución supuesta por la hipótesis nula, pero más posible de ocurrir si la hipótesis alternativa es cierta, constituye evidencia en contra de H_0 y en favor de H_1. Si se establece una probabilidad de cometer el error de rechazar H_0 cuando en verdad p tiene una distribución con los parámetros especificados en dicha hipótesis nula, entonces queda determinado un valor crítico p_c de modo que si se obtiene un valor p mayor que p_c se deberá rechazar H_0. Por otra parte, si el estadístico se distribuye con los parámetros especificados en H_1, queda determinada una probabilidad ß de cometer el error de aceptar H_0. El primero de los errores se conoce como "error de tipo I" y el segundo, como "error de tipo II". La probabilidad 1-ß es la potencia para aceptar la hipótesis alternativa cuando en realidad es cierta.

Como se puede observar en la Figura 1, al aumentar α (área rayada a la derecha de p_c), disminuye ß (área rayada a la izquierda de p_c) y viceversa. Además es posible disminuir ß, y por lo tanto aumentar la potencia, aumentando la diferencia entre los parámetros especificados en ambas hipótesis, es decir entre P_0 y P_1. Otra forma de aumentar la potencia de la prueba consiste en disminuir σ_0^2/N y σ_1^2/N a través del aumento del tamaño de muestra, siendo esta última la forma más frecuentemente utilizada para controlar la probabilidad de detectar el verdadero efecto del tratamiento. Al disminuir las variancias, las distribuciones correspondientes a H_0 y H_1 se concentran más alrededor de la media y disminuyen las áreas superpuestas de ambas distribuciones, de modo que, dado un fijo, ß disminuye y aumenta la potencia de la prueba.

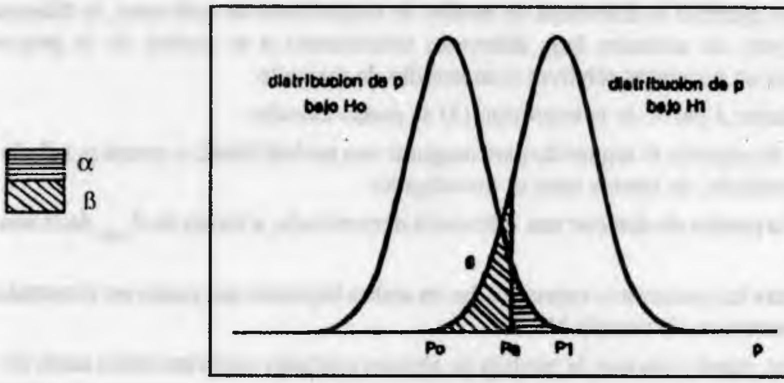


Figura 1: Distribución del estadístico p bajo la hipótesis nula H_0: $P = P_0$, *y bajo la hipótesis alternativa* H_1: $P = P_1$ *con probabilidades de error de tipo I igual a* α *y de error de tipo II igual a* β*, siendo* p_C *el valor crítico del estadístico p.*

Volviendo al ejemplo anterior, para aumentar la probabilidad de detectar una cierta mejora en el porcentaje de preñez debido al nuevo manejo, a valor constante de α, se deberá aumentar el número de animales en el ensayo. El tamaño de muestra requerido para asegurar la potencia deseada, es aquel que, en términos de probabilidad Pr, satisface simultaneamente las igualdades $Pr(p>p_c) = \alpha$, calculada como si H_0 fuera cierta y $Pr(p>p_c) = 1-ß$, calculada como si H_1 fuera cierta.

Se puede expresar el estadístico p en relación a la variable normal estándar como $Z = (p-P_0)(\sigma_0/\sqrt{N})^{-1}$, que, si H_0 es cierta, se distribuye normalmente con media 0 y variancia 1. De esta forma el valor crítico p_c del estadístico p está dado por:

$$p_c = P_0 + Z_{(1-\alpha)} \frac{\sigma_0}{\sqrt{N}} \qquad (1),$$

donde $Z_{(1-\alpha)}$ es el valor de la variable normal estándar correspondiente a una probabilidad acumulada de $1-\alpha$. En caso de pruebas bilaterales se reemplaza $1-\alpha$ por $1-(\alpha/2)$.

Si se expresa p_c de acuerdo con los parámetros especificados en H_1, se tiene que:

$$P_c = P_1 + Z_{(\beta)} \frac{\sigma_1}{\sqrt{N}} \quad (2),$$

donde $Z_{(\beta)}$ es el valor de la variable normal estándar correspondiente a una probabilidad acumulada de ß.

A partir de la diferencia entre (1) y (2) se llega a la siguiente expresión:

$$\sqrt{N}\,|P_1 - P_0| = Z_{(1-\alpha)}\sigma_0 + Z_{(1-\beta)}\sigma_1 \quad (3),$$

donde $Z_{(1-\beta)}$ es el valor de la variable normal estándar correspondiente a una probabilidad acumulada $1-\beta$ y P_1-P_0 es la diferencia entre P_1 y P_0 en valor absoluto.

Esta relación básica puede extenderse a ensayos en los que se plantean hipótesis que involucran otros parámetros, como por ejemplo la diferencia de medias de rendimiento de cultivares, la diferencia en la ganancia media de peso de animales bajo diferentes tratamientos o el cambio de la proporción de respuestas afirmativas en encuestas relativas a un estudio de mercado.

Más específicamente, a partir de la expresión (3) se puede calcular:

- El tamaño total de muestra N requerido para asegurar una probabilidad, o potencia $1-\beta$, de detectar una diferencia determinada, de interés para el investigador.

- La potencia de la prueba de detectar una diferencia determinada, a través de $Z_{(1-\beta)}$ dada una muestra fija de tamaño N.

- La diferencia entre los parámetros especificados en ambas hipótesis que puede ser detectada con una potencia $1-\beta$ si la muestra es de tamaño N.

En ciertos ensayos, puede preverse la pérdida de algunas unidades experimentales antes de finalizar el estudio. Por ejemplo, puede esperarse un cierto porcentaje de muertes en animales, de árboles que se secan o de estacas que no brotan durante el ensayo. Si se espera o se sabe por estudios anteriores que puede perderse una proporción R de unidades experimentales, el tamaño total N_a de muestra requerido puede estimarse de la siguiente forma:

$$N_a = \frac{N}{(1-R)},$$

donde N es el tamaño de muestra calculado sin considerar pérdidas. Con esta corrección se asegura el número de unidades necesarias para el análisis de acuerdo a la potencia prefijada.

De modo similar, es necesario corregir el tamaño de muestra en el cálculo de la potencia y la diferencia mínima a detectar, de manera de no sobreestimar la primera ni subestimar la segunda, reemplazando N por N(1-R) en las fórmulas correspondientes.

En el Apéndice 2 se presenta el diagrama de flujo del programa de computadora para la resolución de los diferentes cálculos de tamaño de muestra, potencia y diferencia mínima verdadera a detectar.

Al probar las hipótesis:

$$H_0: \mu \leqslant 5 \text{ ppm}$$
$$H_1: \mu < 5 \text{ ppm},$$

con un α de 1 por ciento, no se puede concluir, a partir de la muestra, que el promedio del contenido de cobre del campo sea inferior al umbral. Frente a este resultado, interesa conocer cuál es la potencia de la prueba para detectar una diferencia verdadera como la encontrada en la muestra.

De acuerdo a la ecuación (II) presentada en el Apéndice 1 y considerando que no se perdieron muestras de pasto y suponiendo que el desvío estándar de la variable según la hipótesis nula sea igual al que corresponde a la hipótesis alternativa (1,09 ppm), para una diferencia de promedios verdadera de 0,54 ppm (aproximadamente un 11 por ciento del valor umbral), la potencia es de 8,23 por ciento.

Por tal motivo, concluir que no existe deficiencia de cobre en el campo, debido a que el promedio de contenido de este elemento es igual o superior al umbral, podría ser arriesgado. Para lograr una alta probabilidad de detectar la diferencia indicada, por ejemplo una potencia de 80 por ciento, el número de unidades necesarias es 43, de acuerdo a la ecuación (I) para el cálculo de tamaño de muestra presentada en el Apéndice 1.

1.2. Prueba para una proporción poblacional

La variancia σ^2 de una proporción de muestra p queda determinada por la proporción poblacional P de la siguiente forma:

$$\sigma^2 = [P(1-P)]/N.$$

La aplicación de la expresión (3), en este caso, requiere especificar las proporciones poblacionales de las hipótesis nula y alternativa y no solamente la diferencia absoluta entre ambos parámetros, como es posible en la prueba para promedios.

Ejemplo

Se sabe que las variedades nacionales de papa se conservan comúnmente en pilas cubiertas con chala, mientras que las variedades importadas, por su período de reposo más corto, deben ser conservadas en cámaras frigoríficas o silos. Debe tenerse en cuenta, al estimar el rendimiento, que las papas conservadas a la intemperie sufren, en condiciones normales, una merma de alrededor del 18 por ciento (Gaceta Agronómica, 1984). Supóngase que se planea un ensayo para poder decidir si la adopción de la cámara frigorífica para la conservación de las papas nacionales podría reducir la merma a un 10 por ciento, lo que representaría un beneficio económico significativo. Supóngase asimismo, que se desea trabajar con una potencia de 80 por ciento en una prueba unilateral con un de 1 por ciento. Debe entonces fijarse qué tamaño de muestra correspondería tomar si se considera que no existirá pérdida alguna de unidades experimentales por razones ajenas al tratamiento.

De la ecuación (IV) presentada en el Apéndice 1, se desprende que se necesitarían 206 papas para concluir el ensayo, dada la alta probabilidad de éxito prefijada.

APLICACIONES

Se presentan a continuación las aplicaciones de los cálculos de tamaño de muestra y potencia para las pruebas de hipótesis de promedios y proporciones que involucran una sola muestra y dos muestras, incluyendo en este último caso ensayos con observaciones independientes y con observaciones apareadas. El cálculo de la diferencia mínima verdadera a detectar se desarrolla sólo para las pruebas de promedios, debido a que para el caso de proporciones, al ser la diferencia entre parámetros y la variancia funciones de los mismos parámetros, se requiere un procedimiento de tipo iterativo.

En el Apéndice 1 se detallan las ecuaciones para resolver las distintas situaciones planteadas.

1. Ensayos con una sola muestra

Existen ocasiones en que el investigador está interesado en poder detectar un cambio en un promedio o en una proporción teórica o estándar. Este es el caso más sencillo de aplicación de la expresión (3). No obstante, es necesario definir si la hipótesis que se plantea está referida a un promedio o a una proporción.

1.1. Prueba para un promedio poblacional

Las variancias de las poblaciones involucradas en las hipótesis son frecuentemente desconocidas al comienzo del ensayo. En estos casos se pueden emplear las variancias estimadas a partir de una muestra piloto. Es bien conocido el empleo de la distribución *t* de Student en las pruebas de hipótesis para uno o dos promedios cuando las variancias poblacionales son desconocidas. La distribución t se aproxima a la normal estándar a medida que aumenta el tamaño de la muestra, siendo 30 el tamaño requerido para que la aproximación sea adecuada (Johnson y Kotz, 1970). Por esta razón, puede emplearse la expresión (3) para obtener el tamaño de muestra y la potencia aproximados. Sin embargo, es necesario aplicar un factor de corrección para lograr que la aproximación sea más ajustada y evitar, de este modo, subestimar el tamaño de muestra o sobreestimar la potencia. El factor de corrección adecuado, f, (Lachin, 1981) está dado por:

$$f = (gl + 3)/(gl + 1),$$

$$f = (gl + 3)/(gl + 1),$$

donde gl representa los grados de libertad del estadístico *t* correspondiente a la prueba de hipótesis.

El factor de corrección f multiplica al valor de N obtenido en el cálculo de tamaño de muestra y divide al valor de N en el cálculo de la potencia y la diferencia verdadera a detectar.

Ejemplo

En un ensayo realizado para diagnosticar la deficiencia de cobre en producción de vacunos con aptitud carnicera en el partido de Guaminí, se evalúa, como parte del trabajo, el contenido de cobre en pasturas de un establecimiento (Chiatellino, 1992). Con este propósito, se analizan 5 muestras de distintos sectores del campo en estudio. El umbral a partir del cual se considera que existe déficit de cobre en pasturas es de 5 ppm (Quiroga, 1982, citado por Chiatellino, 1992).

2. Ensayos con dos muestras

Frecuentemente, el investigador está interesado en estudiar la efectividad de un tratamiento experimental en comparación con un control o la diferencia en la respuesta de dos tratamientos experimentales. Las hipótesis que se plantean en estos casos involucran a la diferencia de promedios o proporciones poblacionales como parámetro a probar. Generalmente, se establece que este parámetro sea igual a cero en la hipótesis nula y sea un valor distinto de cero en la hipótesis alternativa. Por este motivo, el valor de la diferencia entre los parámetros de ambas hipótesis necesario para la resolución de la expresión (3) es, con frecuencia, el correspondiente al del parámetro definido en la hipótesis alternativa.

2.1. Muestras independientes

En ensayos en los que se comparan dos promedios o proporciones a través de muestras independientes, es conveniente que el tamaño de ambas muestras sea el mismo. Para un número total de unidades muestrales prefijado, la potencia de la prueba se maximiza con muestras de igual tamaño. No obstante, puede ocurrir que al analizar la potencia de un ensayo que no ha sido exitoso, los tamaños de muestra hayan sido, por algún motivo, distintos, o que al planificar un estudio se cuente con un menor número de unidades en alguno de los tratamientos. La desigualdad en el tamaño de las muestras debe quedar reflejada en los cálculos a través de las fracciones de las muestras 1 y 2, Q_1 y Q_2, de modo que la suma de ambas fracciones sea igual a la unidad y el número de unidades para cada muestra sea igual al producto de la fracción correspondiente por el tamaño total de muestra, N.

La extensión de la expresión (3) para el caso de dos muestras independientes es sencilla de interpretar, si se tiene en cuenta que la variancia de una diferencia de variables independientes es igual a la suma de las variancias de dichas variables.

2.1.1. Prueba para promedios poblacionales

Como se señaló en el caso de la prueba para un promedio poblacional, es también frecuente que se desconozcan las variancias poblacionales cuando se comparan dos muestras a través de sus promedios. La estimación de las variancias a partir de las muestras y el supuesto de que las mismas provienen de poblaciones con igual variancia permiten calcular una variancia amalgamada y emplear el estadístico *t*, que seguirá una distribución *t* de Student con $N - 2$ grados de libertad si la hipótesis nula es verdadera.

Ejemplo

Se evaluó el consumo de dos balanceados suministrados *ad libitum* a terneros de la raza Holando-argentino, en los que se deseaba lograr un destete precoz. La composición de ambos balanceados era similar, pero se diferenciaban en el origen, uno era producido en el establecimiento y el otro era un balanceado comercial. La dieta incluía además leche y fardos de alfalfa y pradera para los dos grupos de animales. Cada grupo constaba de 11 terneros de ambos sexos, de aproximadamente un mes de vida. Finalizado el ensayo a los treinta días, se comparó el consumo diario por animal de los balanceados. El balanceado comercial produjo un 8 por ciento más de consumo que el balanceado del establecimiento (1,02 kg/animal-día vs. 0,95 kg/animal-día), diferencia que no resultó significativa en el análisis estadístico. Sin embargo, al calcular la potencia de la prueba para detectar una diferencia verdadera entre promedios como la encontrada en la muestra, mediante la ecuación (VII) del Apéndice 1, considerando la variancia amalgamada del ensayo, 0,108, como estimación de la variancia de las distribuciones correspondientes a H_0 y H_1, con un α del 5 por ciento en prueba bilateral, el resultado arrojó una potencia de 7 por ciento.

2.1.2. Prueba para proporciones poblacionales

En ensayos en los que se comparan dos tratamientos a través de las proporciones poblacionales, a menudo las hipótesis que se plantean son:

$$H_0 : P_1 = P_2.$$
$$H_1 : P_1 \neq P_2$$

La proporción conjunta P de la hipótesis nula se define como un promedio de las dos proporciones poblacionales del siguiente modo:

$$P = Q_1 P_1 + Q_2 P_2.$$

Las variancias de las distribuciones quedan determinadas por los parámetros correspondientes a cada hipótesis, como puede observarse en las ecuaciones (IX) y (X) del Apéndice 1.

Ejemplo

El siguiente ejemplo es parte de un trabajo más extenso llevado a cabo en el Departamento Ojo de Agua, en el sur de la provincia de Santiago del Estero (Ferrari, 1990). Con el objeto de comparar el porcentaje de preñez de dos grupos de animales de distinto estado corporal (categorías 2 y 3 de la escala de Mulvany), se partió de 46 vacas de la categoría más baja y 12 de la otra categoría. Las categorías se determinaron en el momento de la inseminación artificial. No se detectó diferencia significativa en el porcentaje de preñez en la prueba bilateral, a pesar de que la diferencia entre las proporciones de vacas preñadas fue de 0,35 (0,48 para la categoría 2 y 0,83 para la 3). De acuerdo a la ecuación (IX) del Apéndice 1, si se deseara repetir el ensayo, serían necesarios 28 animales en cada grupo para detectar una diferencia como la observada en la muestra con una potencia de 80 por ciento, considerando un α de 5 por ciento.

2.2. Muestras apareadas

En algunos ensayos, las observaciones de ambas muestras están apareadas, en el sentido que cada observación de una muestra se asocia con determinada observación de la otra. Puede resultar apareamiento, por ejemplo, por grado de fertilidad del suelo, por sexo o edad de animales, o porque las mismas unidades se observan antes y después de cierto tratamiento.

Se consideran a continuación los ensayos referidos a promedios o proporciones con muestras apareadas. En ambos casos, los tamaños de muestra son necesariamente iguales.

2.2.1. Prueba para promedios poblacionales

La variancia de una diferencia de variables no es igual, en este caso, a la suma de las variancias individuales de las variables, como en los ensayos con observaciones independientes. Debe aquí considerarse la correlación de las unidades apareadas, de modo que:

$$\sigma^2_d = 2\sigma^2(1-\rho),$$

donde σ^2_d es la variancia de las diferencias de las observaciones apareadas, σ^2 es la variancia de las observaciones individuales y ρ es el coeficiente de correlación de los pares.

Se puede contar con alguna estimación de σ^2_d de algún estudio previo o piloto. Si no es así y se posee información de σ^2, pero no de ρ, es conveniente asumir un valor bajo (positivo) de ρ, de modo de no sobreestimar la potencia ni subestimar el tamaño de muestra.

Ejemplo

En un ensayo para comparar el rendimiento de tomate "cereza" bajo dos sistemas de cultivo, conducción a un tallo y a cuatro tallos, se diseñaron bloques considerando las distintas distancias de las parcelas a la entrada del invernáculo (Buzzi, 1992). Como no se poseía información previa, se llevó a cabo el ensayo a manera de estudio piloto, con sólo cuatro pares de parcelas, debido a la escasa superficie disponible. El análisis estadístico no evidenció diferencias significativas. Si al investigador le interesara detectar una diferencia de 3,5 kg por parcela en el rendimiento total de tomate, con una potencia de 80 por ciento, considerando al desvío estándar de las diferencias obtenido en el ensayo anterior, 3,70 kg, como una estimación del desvío poblacional, el número de pares de parcelas necesario para el próximo ensayo debería ser de 11, con α de 5 por ciento en una prueba bilateral, de acuerdo a la ecuación XI del Apéndice 1.

2.2.2. Prueba para proporciones poblacionales

En la Tabla 1 se presentan las observaciones que resultan de ensayos en los que se comparan proporciones con muestras apareadas. Las situaciones a y b que aparecen en la tabla pueden ser momentos, antes y después de un tratamiento, o dos tratamientos en un ensayo diseñado en bloques. Como ejemplo de este último caso, se puede considerar la comparación de dos medios de cultivo a partir de yemas brotadas de tejidos, en un diseño en el que se establecen bloques.

Si se simboliza con un signo + a las yemas brotadas y con un signo - a las no brotadas, se observa que en una proporción p_{+} de bloques brotaron yemas en el medio de cultivo a y en una proporción $p_{\cdot+}$ de bloques brotaron yemas en el medio de cultivo b. Las proporciones p_{+-} y p_{-+} representan las proporciones de bloques en que brotaron las yemas en el medio de cultivo a, pero no en el b, y viceversa, respectivamente.

Tabla 1: Proporción de respuestas a dos situaciones clasificadas en dos categorías (+ y -).

		Situación b +	Situación b -	Total
Situación a	+	p_{++}	p_{+-}	$p_{+\cdot}$
	-	p_{-+}	p_{--}	$p_{-\cdot}$
Total		$p_{\cdot+}$	$p_{\cdot-}$	1

El punto en las proporciones marginales indica la suma de las proporciones a través del correspondiente subíndice.

La hipótesis nula que se plantea en este tipo de estudio es la siguiente:

$$H_0: P_{+.} = P_{.+}$$

La diferencia entre $P_{+.}$ y $P_{.+}$ está dada por las proporciones P_{+-} y P_{-+}. Por este motivo la hipótesis nula puede establecerse también en términos de las proporciones discordantes como:

$$H_0: P_{+-} = P_{-+}$$

La variancia de la diferencia de las proporciones discordantes estimadas según esta hipótesis, de acuerdo a Lachin (1981), está dada por:

$$\sigma_0^2 = 2P,$$

donde $P = 1/2\,(P_{-+} + P_{+-})$.

La variancia de la distribución correspondiente a la hipótesis alternativa se define como:

$$\sigma_1^2 = (2\,P_{-+}\,P_{+-})/P\ .$$

Para calcular el tamaño de muestra y la potencia se debe determinar cuál es la diferencia de las proporciones discordantes de interés para el investigador.

Ejemplo

Supóngase que un laboratorio de productos veterinarios desea lanzar una publicidad al mercado para incrementar las ventas de su antiparasitario frente a otro considerado líder en su género. La empresa, antes de decidir el lanzamiento, desea probar si es posible lograr aumentos del 20 por ciento en su volumen de ventas debido a tal publicidad. Por ello se decide tomar una muestra de consumidores de antiparasitarios antes y después de una prueba piloto de la publicidad y verificar la decisión de compra del producto en cuestión. El número de personas encuestadas antes y después de la publicidad dependerá de las proporciones de los que antes utilizaban el antiparasitario de la competencia y luego cambiaron por el del laboratorio y de los que cambiaron en sentido inverso. Si se conoce que la proporción de cambio de marca, P_{+-}, en ese rubro es aproximadamente de 0,10, un 20 por ciento más a favor del antiparasitario de interés implica una proporción P-+ de 0,30. De este modo, la resolución de la ecuación (XIV) del Apéndice 1, para una potencia de 85 por ciento, con una probabilidad de error de tipo I de 2,5 por ciento para una prueba unilateral, y considerando como posible una proporción de 10 por ciento de no respuestas a la encuesta, indica que deberán incluirse 91 personas en la muestra.

Es interesante observar que para una misma diferencia entre proporciones, al variar las mismas, varía también el tamaño de muestra. En la Tabla 2 se presentan algunas estructuras posibles de proporciones que resultan en distintos tamaños de muestra, sin considerar unidades perdidas.

Tabla 2: Tamaños de muestra para detectar un 20, 15 y 10 por ciento de diferencia en las proporciones discordantes, con una potencia de 85 por ciento y un α de 2,5 por ciento para una prueba unilateral, según distintas proporciones.

P-+	P+-	Diferencia	N
0,40	0,20	0,20	130
0,30	0,10	0,20	82
0,25	0,05	0,20	56
0,35	0,20	0,15	214
0,25	0,10	0,15	131
0,20	0,05	0,15	87
0,30	0,20	0,10	443
0,20	0,10	0,10	259
0,15	0,05	0,10	164

2.3. Muestras independientes con observaciones apareadas

En ciertas ocasiones el investigador está interesado en comparar dos tratamientos a través de muestras independientes en las que se han hecho observaciones en dos momentos, antes y después de aplicado el tratamiento. En este caso, las muestras pueden ser de distinto tamaño.

2.3.1. Prueba para promedios poblacionales

Son frecuentes los ensayos en los que se comparan promedios de ganancia de peso de animales a partir del consumo de distintas raciones o de aumento de rendimiento de frutales luego de la aplicación de fertilizantes u otros productos agroquímicos. El análisis de estos ensayos es muy similar al ya indicado para la comparación de promedios en diseños con muestras independientes. La variable respuesta es, en este caso, la diferencia entre las observaciones antes y después de aplicado el tratamiento. Por esta razón, las ecuaciones necesarias para resolver el tamaño de muestra, la potencia y la diferencia mínima verdadera a detectar son las mismas que las ya vistas en el punto 2.1.1. (ecuaciones VI, VII y VIII del Apéndice 1).

Ejemplo

Supóngase que un productor ganadero pone a disposición de un ingeniero agrónomo 200 animales para estudiar el efecto de un anabólico en la ganancia de peso de vacunos. Los animales, que son asignados al azar al grupo control o al tratado con el anabólico, se pesan al principio del ensayo y a los dos meses siguientes, considerándose a la diferencia de peso como variable respuesta. Se cuenta con información de un ensayo previo similar, en el que el desvío estándar estimado de la ganancia de peso fue de 9 kg. Con un tamaño de muestra de 100 en cada grupo ($Q_1 = Q_2 = 0{,}5$) y de acuerdo a la ecuación (VIII) del Apéndice 1, se espera poder detectar una diferencia de ganancia de peso verdadera de 3,83 kg con una potencia del 85 por ciento, considerando una prueba bilateral con α de 5 por ciento. Se plantea aquí si detectar una diferencia tan pequeña puede ser relevante en la práctica como para justificar el costo de un ensayo que involucra el manejo de un número tan grande de animales. Un ensayo con una muestra de menor tamaño y con una alta potencia puede llegar a servir para detectar una diferencia de significación para el productor ganadero.

2.3.2. Prueba para proporciones poblacionales

Una extensión de la prueba para proporciones poblacionales con muestras apareadas es la que considera dos muestras independientes con observaciones apareadas.

Un investigador puede estar interesado en comparar las proporciones de cambio de respuesta de dos tratamientos. No interesa, en este caso, si las proporciones discordantes (P_{+-} y P_{-+}) son iguales en cada muestra, sino si la diferencia entre ambas proporciones es la misma en los dos tratamientos. De este modo, si se denota con δ_1 y δ_2 a la diferencia entre las proporciones discordantes de las muestras 1 y 2, respectivamente, la hipótesis nula puede establecerse del siguiente modo:

$$H_0: \delta_1 - \delta_2 = 0$$

Esta hipótesis plantea la ausencia de interacción de los tratamientos con los momentos de observación. El estadístico para probar H_0 está dado por:

$$d_1 - d_2 = p_{2-+} - p_{2+-} - p_{1-+} + p_{1+-},$$

donde d_1 y d_2 estiman a δ_1 y δ_2 respectivamente.

La hipótesis de ausencia de interacción puede implicar que los dos grupos tengan las mismas proporciones de cambio en ambos sentidos, de modo que se pueden establecer parámetros comunes, P_{+-} y P_{-+}, de la siguiente forma:

$$P_{+-} = Q_1 * P_{1+-} + Q_2 * P_{2+-},$$
$$P_{-+} = Q_1 * P_{1-+} + Q_2 * P_{2-+}.$$

Sin embargo, se pueden establecer modelos alternativos de ausencia de interacción, que no involucren un supuesto tan restrictivo. La definición de las proporciones, de acuerdo a estos modelos alternativos, se presentan en el Apéndice 1, como así también las variancias correspondientes a las distribuciones de H_0 y H_1. Las ecuaciones (XVI) y (XVII) del mismo apéndice se emplean para los cálculos de tamaños de muestra y potencia estadística.

Ejemplo

Supóngase que un investigador está interesado en estudiar el efecto de un nuevo producto medicinal para el tratamiento de cierta enfermedad ocular en bovinos. Cuenta, para ello, con un gran número de animales de un establecimiento donde existe la infección, la que, indefectiblemente, provoca la pérdida del ojo si no es tratada, por lo que debe incorporar el producto estándar en la comparación a modo de control. Para calcular el tamaño de muestra necesario para el ensayo, el investigador deberá plantearse qué diferencia en las proporciones discordantes entre tratamientos desea poder detectar con una alta probabilidad. Si está interesado en que esta diferencia sea de 0,20 y sabe por ensayos anteriores que una proporción de 0,10 animales sanos se enferman en lotes infectados aunque hayan recibido medicación ($P_{1+-} = P_{2+-} = 0,10$), y que la proporción de los que se curan con el producto estándar es de 0,30, con un de 2,5 por ciento para una prueba unilateral y una potencia de 90 por ciento se requieren 335 animales en el ensayo. Si se espera una pérdida del 2 por ciento de animales, se deberá aumentar el número de animales a 171 por grupo tratado.

DISCUSION

La determinación del tamaño de muestra al comienzo de cualquier ensayo es una forma de asegurar una alta chance de éxito del mismo.

Frecuentemente el estudio de la potencia de la prueba es posterior a un resultado estadísticamente insignificante. Si la potencia es baja, el ensayo es inconclusivo debido a que la aceptación de H_0 no implica necesariamente la equivalencia de los tratamientos dada la alta probabilidad de cometer un error de tipo II. Si, por ejemplo, no se detecta una diferencia moderada en el porcentaje de germinación de semillas de un cereal, tratadas y no tratadas con un fungicida para el control del carbón volador en un ensayo con potencia menor a 70 por ciento, la probabilidad de que el ensayo haya fallado en detectar la verdadera diferencia es alta, de modo que es arriesgado considerar que el fungicida no afecta el poder germinativo de la semilla.

El cálculo de la diferencia a detectar puede ser relevante tanto en la etapa de planificación como en la posterior al ensayo. Una diferencia muy pequeña a los fines prácticos puede indicar que el tamaño de muestra planeado en el diseño es excesivamente grande. Es posible, por ejemplo, que detectar un aumento de 10 kg/ha en el rendimiento de maíz por el empleo de un fertilizante involucre un alto costo y no justifique la inversión realizada en el ensayo por no implicar sus resultados un beneficio potencial suficiente para el productor. En este caso, la utilización de un menor número de unidades experimentales permitiría detectar con una alta probabilidad una diferencia en el rendimiento de maíz sólo cuando ésta fuera relevante para los objetivos desde la perspectiva del productor.

En ciertas ocasiones, el investigador está dispuesto a planificar su ensayo correctamente, pero no cuenta con el suficiente material experimental o superficie física para realizar el mismo con el número de unidades adecuado. En estos casos, es conveniente analizar cuidadosamente cuál es la diferencia máxima entre tratamientos que resultaría interesante detectar, sin disminuir la potencia. El investigador debe decidir si comenzar un ensayo con escasos recursos y alta probabilidad de fracaso o esperar a contar con las condiciones necesarias para asegurar una alta chance de éxito para alcanzar los objetivos propuestos.

Reconocimiento

Los autores agradecen a la Lic. María del Carmen Fabrizio por la lectura crítica del manuscrito preliminar.

NOTA: El programa mencionado se encuentra disponible en la Cátedra de Estadística.

BIBLIOGRAFIA

-**ABRAMOWITZ, M.** and **STEGUN, I.**, 1964. Handbook of mathematical functions. National Bureau of Standards. Washington, D.C.

-**BUZZI, A. M.**, 1992. Evaluación del comportamiento de una variedad "middle" de tomate cereza *(Licopersicum esculentum* var. *ceraciforme)* con dos tipos de conducción, en producción tardía bajo invernáculo. Inédito, trabajo de intensificación para optar al título de Ingeniero Agrónomo, Facultad de Agronomía, Universidad de Buenos Aires.

-**COHEN, J.**, 1977. Statistical power analysis for the behavioral sciences. Academic Press, New York.

-**CHIATELLINO, D. L.**, 1992. Determinación de deficiencia y tratamiento de cobre. Inédito, trabajo de intensificación para optar al título de Ingeniero Agrónomo, Facultad de Agronomía, Universidad de Buenos Aires.

-**CHOU, Y. L.**, 1986. Análisis estadístico. Editorial Interamericana, México, 3 Ed., pp. 224-225.

- **GACETA AGRONOMICA** 1984. El almacenaje de la papa, 1984. *Gaceta Agronómica.* 4 (17): 66-67.

-**FERRARI, T. G.**, 1990. Estado corporal y su influencia sobre la aptitud reproductiva de hembras de raza Brahman. Inédito, trabajo de intensificación para optar al título de Ingeniero Agrónomo, Facultad de Agronomía, Universidad de Buenos Aires.

-**JOHNSON, N. L.** y **KOTZ, S.**, 1970. Distributions in Statistics: continuous univariate distributions. Wiley, New York, Vol. 2, p. 101.

-**LACHIN, J. M.**, 1981. Introduction to sample size determination and power analysis for clinical trials. Controlled *Clinical Trials*, 2: 93-113.

-**QUICKBASIC** versión 4.5, 1985-1988. Microsoft Corporation.

-**QUIROGA, M. A.**, 1983. Hipocuremia experimental en bovinos. *Gaceta Veterinaria*, 15: 383.

-**SNEDECOR, G. W.** y **COCHRAN, W. G.**, 1979. Métodos estadísticos. CECSA, México, 6 Ed., pp. 147-151, 278-280.

STEEL, R. G. y **TORRIE, J. H.**, 1980. Principles and procedures of statistics. McGraw-Hill, New York, 2 Ed., pp. 113-119.

APENDICE 1

Ecuaciones para el cálculo de tamaño de muestra, potencia y diferencia de parámetros.

1. Ensayos con una sola muestra

1.1. Prueba para un promedio poblacional

Tamaño de muestra

$$N=\left[\frac{Z_{(1-\alpha)}\sigma_0+Z_{(1-\beta)}\sigma_1}{\mu_1-\mu_0}\right]^2 \qquad (I)\ ,$$

donde:

N es el tamaño de muestra, $Z_{(1-\alpha)}$ es el valor de la variable normal estándar correspondiente a una probabilidad acumulada $1-\alpha$, $Z_{(1-\beta)}$ es el valor de la variable normal estándar correspondiente a una probabilidad acumulada $1-\beta$ y $\mu_1-\mu_0$ es la diferencia de los parámetros planteados en H_1 y H_0, respectivamente. σ_0 y σ_1 son las variancias poblacionales.

Si no se conoce la variancia poblacional y se emplea el estadístico *t* de Student en la prueba de hipótesis, considerando una proporción R de unidades perdidas, el tamaño de muestra debe ser corregido del siguiente modo:

$$N_a = Nf/(1-R)\ ,$$

donde:

N_a es el tamaño de muestra ajustado, $f = [(gl) + 3]/[(gl) + 1]$, gl son los grados de libertad asociados al estadístico *t*, en este caso gl = N-1.

Potencia

$$Z_{(1-\beta)}=\frac{\sqrt{\frac{N(1-R)}{f}}|\mu_1-\mu_0|-Z_{(1-\alpha)}\sigma_0}{\sigma_1} \qquad (II)\ ,$$

Diferencia de parámetros

$$|\mu_1-\mu_0|=\frac{Z_{(1-\alpha)}\sigma_0+Z_{(1-\beta)}\sigma_1}{\sqrt{\frac{N(1-R)}{f}}} \quad (III)\ ,$$

1.2. Prueba para una proporción poblacional

Tamaño de muestra

$$N=\left[\frac{Z_{(1-\alpha)}\sqrt{P_0(1-P_0)}+Z_{(1-\beta)}\sqrt{P_1(1-P_1)}}{P_1-P_0}\right]^2 \quad (IV)\ ,$$

donde P_0 es la proporción poblacional de la distribución correspondiente a H_0 y P_1 es la proporción poblacional de la distribución correspondiente a H_1.

El tamaño de la muestra ajustado está dado por:

$$N_a = N/(1-R)$$

Potencia

$$Z_{(1-\beta)}=\frac{\sqrt{N(1-R)}|P_1-P_0|-Z_{(1-\alpha)}\sqrt{P_0(1-P_0)}}{\sqrt{P_1(1-P_1)}} \quad (V)\ .$$

2. Ensayos con dos muestras

2.1. Muestras independientes

2.1.1. Prueba para promedios poblacionales

Tamaño de muestra

$$N=\frac{\sigma^2(Q_2^{-1}+Q_1^{-1})(Z_{(1-\alpha)}+Z_{(1-\beta)})^2}{(\mu_1-\mu_2)^2} \quad (VI)\ ,$$

donde:

σ^2 es la variancia común a las distribuciones correspondientes a H_0 y H_1, Q_1 es la fracción del total de la muestra correspondiente al grupo 1, Q_2 es la fracción del total de la muestra correspondiente al grupo 2 y μ_1 y μ_2 son los promedios poblacionales de los grupos $\mu 1$ y $\mu 2$, respectivamente.

El tamaño de muestra ajustado está dado por:

$$N_a = Nf/(1-R)$$

En el cálculo de f, en este caso, gl = N-2.

Potencia

$$Z_{(1-\beta)}=\frac{|\mu_1-\mu_2|\sqrt{\frac{N(1-R)}{f}}-Z_{(1-\alpha)}\sigma\sqrt{Q_2^{-1}+Q_1^{-1}}}{\sigma\sqrt{Q_2^{-1}+Q_1^{-1}}} \qquad (VII) \quad .$$

Diferencia de parámetros

$$|\mu_1-\mu_2|=\frac{\sigma\sqrt{Q_2^{-1}+Q_1^{-1}}}{\sqrt{\frac{N(1-R)}{f}}}\left(Z_{(1-\alpha)}+Z_{(1-\beta)}\right) \qquad (VIII) \quad .$$

2.1.2. Prueba para proporciones poblacionales

Tamaño de muestra

$$N=\left[\frac{Z_{(1-\alpha)}\sigma_0+Z_{(1-\beta)}\sigma_1}{|P_1-P_2|}\right]^2 \qquad (IX) \quad ,$$

donde:

$\sigma_0 = [P(1-P)(Q_2^{-1}+Q_1^{-1})]^{1/2}$, siendo:

$P = Q_2P_2+Q_1P_1$, P1 y P2 son las proporciones poblacionales de las muestras 1 y 2, respectivamente, y

$\sigma_1 = [P_2(1-P_2)Q_2^{-1}+P_1(1-P_1)Q_1^{-1}]^{1/2}$.

El tamaño de muestra ajustado está dado por:

$N_a = N/(1-R)$.

Potencia

$$Z_{(1-\beta)}=\frac{|P_2-P_1|\sqrt{N(1-R)}-Z_{(1-\alpha)}\sigma_0}{\sigma_1} \qquad (X) \quad .$$

2.2. Muestras apareadas

2.2.1. Prueba para promedios poblacionales

Tamaño de muestra

$$N=\frac{(Z_{(1-\alpha)}+Z_{(1-\beta)})^2\sigma_d^2}{(\mu_1-\mu_2)^2} \qquad (XI) \quad ,$$

donde:

N es el número de pares en la muestra y σ_d^2 es la variancia de las diferencias de las observaciones apareadas.

El tamaño de muestra ajustado está dado por:

$$N_a = Nf/(1\text{-}R)$$

En el cálculo de f, gl = N-1.

Potencia

$$Z_{(1-\beta)} = \frac{|\mu_1-\mu_2|\sqrt{\frac{N(1-R)}{f}} - Z_{(1-\alpha)}\sigma_d}{\sigma_d} \qquad (XII) \quad .$$

Diferencia de parámetros

$$|\mu_1-\mu_2| = \frac{\sigma_d}{\sqrt{\frac{N(1-R)}{f}}}\left(Z_{(1-\alpha)}+Z_{(1-\beta)}\right) \qquad (XIII) \quad .$$

2.2.2. Prueba para proporciones poblacionales

Tamaño de muestra

$$N = \left[\frac{Z_{(1-\alpha)}\sqrt{2P}+Z_{(1-\beta)}\sqrt{\frac{2P_{-+}P_{+-}}{P}}}{P_{-+}-P_{+-}}\right]^2 \qquad (XIV) \quad ,$$

donde:

P_{-+} es la proporción de cambio a respuesta positiva, P_{+-} es la proporción de cambio a respuesta negativa y $P = 1/2\,(P_{-+} + P_{+-})$.

El tamaño de muestra ajustado está dado por:

$N_a = N/(1\text{-}R)$.

Potencia

$$Z_{(1-\beta)} = \frac{\sqrt{N(1-R)}\,|P_{-+}-P_{+-}| - Z_{(1-\alpha)}\sqrt{2P}}{\sqrt{\frac{2P_{-+}P_{+-}}{P}}} \qquad (XV) \quad .$$

2.3. Muestras independientes con observaciones apareadas

2.3.1. Prueba para promedios poblacionales

Las ecuaciones para calcular N, $Z_{(1-\beta)}$ y μ_1-μ_2 son las mismas que para la prueba para promedios poblacionales con muestras independientes. La variable respuesta, en este caso, es la diferencia entre las observaciones de dos momentos, antes y después.

2.3.2. Prueba para proporciones poblacionales

Tamaño de muestra

$$N=\left[\frac{Z_{(1-\alpha)}\sigma_0+Z_{(1-\beta)}\sigma_1}{P_{2-+}-P_{2+-}-P_{1-+}+P_{1+-}}\right]^2 \qquad (XVI)\ ,$$

donde:

$$\sigma_0=\sqrt{\frac{4P'_{2-+}P'_{2+-}}{Q_2\,(P'_{2-+}+P'_{2+-})}+\frac{4P'_{1-+}P'_{1+-}}{Q_1\,(P'_{1-+}+P'_{1+-})}}\ ,$$

$$\sigma_1=\sqrt{\frac{4P_{2-+}P_{2+-}}{Q_2\,(P_{2-+}+P_{2+-})}+\frac{4P_{1-+}P_{1+-}}{Q_1\,(P_{1-+}+P_{1+-})}}\ ,$$

P_{i-+} y P_{i+-} son las proporciones esperadas de cambio de respuesta de la población i, (i = 1, 2).
$P'_{1+-} = P_{1+-}$, $P'_{1-+} = P_{1-+} + 1/2\,(P_{2-+} - P_{2+-} - P_{1-+} + P_{1+-})$, $P'_{2+-} = P_{2+-}$, $P'_{2-+} = P_{2-+} - 1/2\,(P_{2-+} - P_{2+-} - P_{1-+} + P_{1+-})$.

El tamaño de muestra ajustado está dado por:

$N_a = N/(1-R)$.

Potencia

$$Z_{(1-\beta)}=\frac{\sqrt{N(1-R)}\,|P_{2-+}-P_{2+-}-P_{1-+}+P_{1+-}|-Z_{(1-\alpha)}\sigma_0}{\sigma_1} \qquad (XVII)\ .$$

APENDICE 2

El programa de computación está escrito en lenguaje *Quick Basic* (1985-1988).

El valor del nivel de significación utilizado en las fórmulas se corresponde con hipótesis alternativas unilaterales. En el caso de plantearse hipótesis alternativas bilaterales el α a ingresar en el programa debería ser el del test dividido por 2.

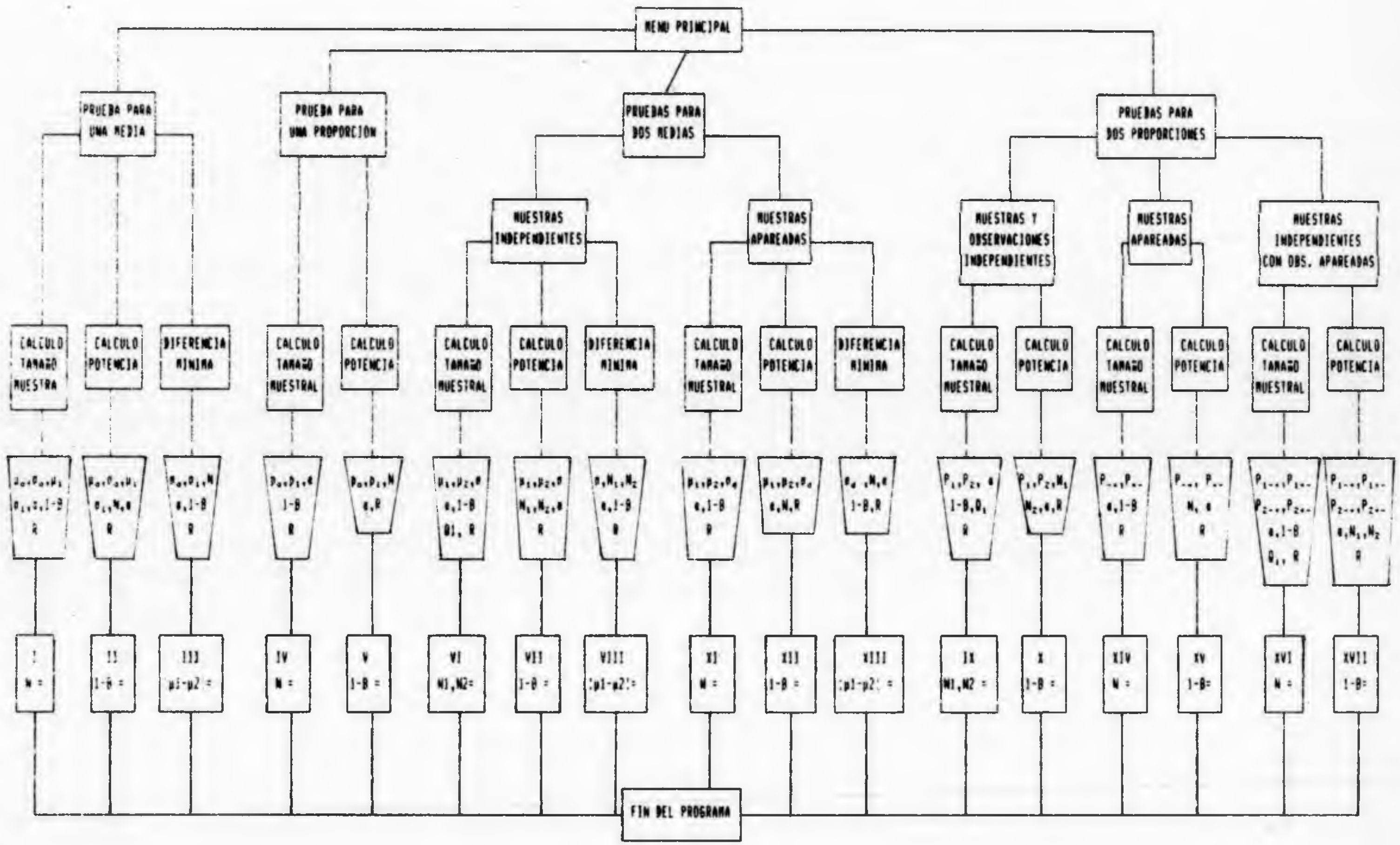
Para el cálculo del tamaño de muestra el programa aproxima al entero superior. Por ejemplo, si el valor de N calculado es de 461,02, el resultado presentado en pantalla es 462.

El rango de valores admitido para α está entre 0,001 y 0,10 y para 1-ß (cuando es parte del *"input"*) entre 0,50 y 0,99.

Para el cálculo de la potencia en las pruebas para una y dos proporciones el tamaño muestral total debe ser mayor o igual a 30 para asegurar una buena aproximación a la distribución normal.

En la Figura 2 se presenta el diagrama de flujo del programa. El desarrollo completo del mismo se encuentra en el archivo de la Cátedra de Estadística de la Facultad de Agronomía de la Universidad de Buenos Aires, disponible para cualquier consulta.

Figura 2: Diagrama de flujo del programa para calcular tamaño de muestra, potencia y diferencia mínima verdadera



Nota: -Todos los símbolos siguen la notación del texto.
- Los números romanos corresponden a las ecuaciones del apéndice.