

INFORMACIÓN TÉCNICA

POTENCIA DE PRUEBA: LA GRAN AUSENTE EN MUCHOS TRABAJOS CIENTÍFICOS¹

María Isabel González-Lutz²

RESUMEN

Potencia de prueba: la gran ausente en muchos trabajos científicos. Se pretende destacar la importancia de la consideración del error tipo II y por ende la potencia de la prueba en la investigación científica, particularmente en el campo del diseño experimental. Se presenta su uso con base en un ejemplo hipotético y en un ejemplo real de un experimento en el campo agronómico en el que se estudió el efecto de la concentración de fertilizante sobre la vida útil y la calidad de las flores de *Phalaenopsis* y se concluyó, sustentando en una potencia de prueba alta, que aplicando fertilizante foliar en baja concentración es posible aumentar la calidad de las flores sin detrimento de su vida útil.

Palabras clave: Hipótesis, nivel de significancia, error tipo II, experimentos, *Phalaenopsis*.

ABSTRACT

Power of the test: the big absentee in many scientific works. The objective of this work is to highlight the importance of considering type II error and therefore the power of the test in scientific research, particularly in experimental designs. Its use and determination are illustrated with both hypothetical and real examples in the agricultural field. In the latter, the effect of fertilizer concentration on quality and durability of *Phalaenopsis* flowers was studied. The conclusion, based on the high power of the test, was that a low concentration of fertilizer increases the quality of the flower, not affecting the durability.

Key words: Hypothesis, significance level, type II error, experiments, *Phalaenopsis*.



INTRODUCCIÓN

Cuando se prueba una hipótesis hay dos posibles resultados: rechazarla o no rechazarla (los estadísticos solemos decir que no rechazar una hipótesis no implica aceptarla, por lo que utilizamos la frase “no hay evidencia para rechazar la hipótesis nula”). Cualquiera que sea la decisión hay una probabilidad conocida de cometer un error. Cuando la hipótesis nula se rechaza, esa probabilidad es el error tipo I al que se ha llamado tradicionalmente “nivel de significancia”. Cuando la

hipótesis nula no se rechaza la probabilidad de equivocarse se denomina error tipo II. La consideración del primero está presente en todas las investigaciones en las que se realiza prueba de hipótesis, la consideración del segundo suele estar ausente.

El objetivo de este trabajo es destacar la importancia del error tipo II, cómo éste se manifiesta en la potencia de la prueba e ilustrar su cálculo y su interpretación, tanto cuando la hipótesis que se desea probar está planteada en la hipótesis nula como cuando no lo está.

¹ Recibido: 24 de mayo, 2007. Aceptado: 26 de agosto, 2008.

² Escuela de Estadística, Universidad de Costa Rica (UCR), San Pedro, San José, Costa Rica. migonzal@cariari.ucr.ac.cr

Las hipótesis y su planteamiento

Ha sido tradicional en la investigación científica que las hipótesis planteadas en el campo específico de una investigación se conviertan, cuando se plantean en términos estadísticos, en una “hipótesis nula” que se desea rechazar. En términos generales, el rechazo de la hipótesis nula lleva a la aceptación de la hipótesis alternativa, que es, usualmente, la hipótesis sustantiva que se desea probar. Para citar algún ejemplo, Sierra *et al.* (2006) plantearon como objetivo “evaluar la asociación entre la presencia de gastritis atrófica en el antro con la presencia de infección por *H. pylori*”, este objetivo se convirtió en la hipótesis nula “el coeficiente de regresión asociado a *H. pylori* es cero” cuyo rechazo implica aceptar la presencia de una asociación. Saborío *et al.* (1994) plantearon como objetivo “evaluar la factibilidad de reducir la contaminación de *Lactuca sativa* por *Vibrio cholerae* mediante la aplicación poscosecha de productos desinfectantes” la cual produjo, entre otras, la hipótesis nula “un promedio de bacterias de 10^6 inoculado en el vegetal no cambia cuando se aplica Kilol en una concentración de 1.000 ppm” cuyo rechazo indica que el Kilol puede utilizarse como agente bactericida.

No siempre las hipótesis sustantivas llevan a una hipótesis nula que se desea rechazar, muchas veces lo que se busca es precisamente lo contrario. Como ejemplo de ésto, Hernández y Herrera (2005) plantearon como objetivo “seleccionar herbicidas promisorios en condiciones de invernadero que muestren selectividad a la especie de pasto *B. brizantha* cv. Diamantes 1” el cual se tradujo, entre otras, a la hipótesis nula “la biomasa promedio en el testigo es igual a la biomasa promedio del Bispiribac sodio”. Si no es posible rechazar esta hipótesis el Bispiribac sodio es un herbicida que muestra selectividad en invernadero. Otro ejemplo es el de Wong *et al.* (1997) que buscó demostrar que la tasa de filtración de las pianguas no se altera cuando se cultivan en medios ácidos no tóxicos.

El nivel de significancia y la potencia de la prueba. Un ejemplo hipotético

Cuando se prueba una hipótesis, como ya se mencionó, hay dos posibles decisiones: aceptarla o rechazarla y la decisión que se tome puede ser correcta

o incorrecta. Esto se presenta de la siguiente manera, tal como se observa en múltiples textos de Estadística (ver, por ejemplo, Senders 1958, Daniel 2002, Levin 1988):

Supongamos una prueba así:

$$H_0: \mu = 36$$

$$H_1: \mu > 36$$

En el Cuadro 1 se observan las posibles decisiones para la prueba de la hipótesis nula y las probabilidades asociadas a cada una de ellas.

Cuadro 1. Decisiones con respecto a la hipótesis nula y probabilidades asociadas a cada una de ellas.

	Decisión:	
	Rechazar H_0	Aceptar H_0
H_0 cierta	Error tipo I α = nivel de significancia	Decisión correcta 1- α
H_0 falsa	Decisión correcta β = POTENCIA DE LA PRUEBA	Error tipo II β

Recuérdese que la hipótesis que se prueba y para la cual tomamos la decisión es H_0 la hipótesis nula. Si la hipótesis nula es falsa eso implica automáticamente que la alternativa es la cierta. En la prueba que estamos realizando, esto indica que el promedio verdadero no sería 36 sino uno mayor. Supongamos que en efecto el promedio verdadero no es 36, la hipótesis nula es falsa y esto implica la existencia de una diferencia en el promedio, de cierto tamaño o mayor, que nuestra prueba debería ser capaz de detectar. La probabilidad de que nuestra prueba detecte esa diferencia es la potencia de la prueba. En la Figura 1 (que se elaboró utilizando el módulo “Power animation” del paquete estadístico JMP, versión 4) se muestran el error Tipo I (α), el error Tipo II (β) y la potencia de prueba (1- β). El valor de 38,4 es el límite entre la zona de aceptación y la zona de rechazo para la hipótesis nula: valores inferiores del promedio muestral llevan a la aceptación,

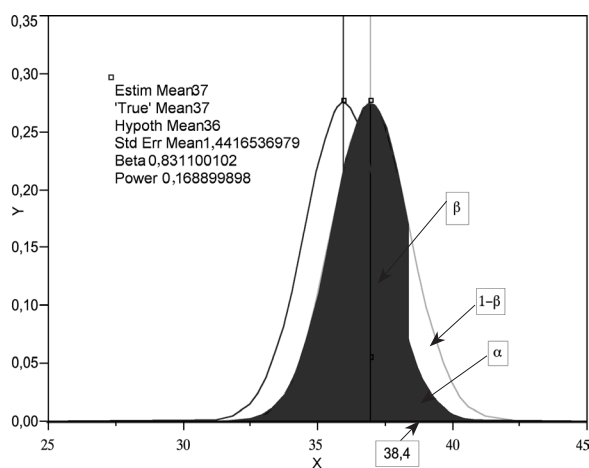


Figura 1. Error Tipo I, error Tipo II y potencia de prueba en una prueba de hipótesis para un promedio.

valores superiores al rechazo. Supongamos que la media verdadera es 37, el cálculo es, entonces,

$$1-\beta = P(X > 38,4 / \mu = 37) = 0,168$$

Supongamos, también, que en la muestra de tamaño 65 que se tomó para probar la hipótesis se obtuvo un promedio de 38 con el cual se decide no rechazar la hipótesis nula. El error es consecuencia de una potencia de prueba baja: la prueba tiene una probabilidad de 0,168 de detectar una diferencia de tamaño 1 lo que implica que sólo el 16,8% de las pruebas que se hagan en estas condiciones serán capaces de detectar esa diferencia y llevarán a una conclusión correcta.

La manera de aumentar esa potencia de prueba es aumentando el tamaño de la muestra. En la Figura 2 se observa el resultado de aumentar el tamaño de la muestra a 130. Nótese que la precisión aumentó (el error estándar del promedio es ahora de 0,73 en vez de 1,44) y la potencia de prueba se elevó a 0,38. Evidentemente, este resultado podría ser mejor o no ser tan bueno, dependiendo de la muestra que se obtenga por azar.

El nivel de significancia y la potencia de la prueba. Un ejemplo real

Se realizó una investigación titulada “Determinación del efecto de la concentración de fertilizante sobre

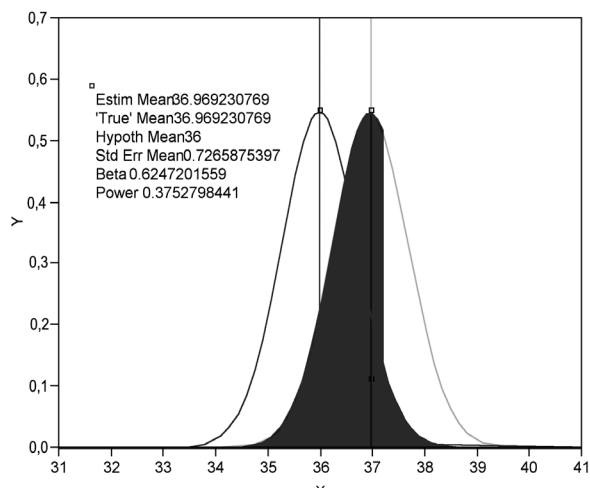


Figura 2. Modificaciones en el Error Tipo I, error Tipo II y potencia de prueba en una prueba de hipótesis para un promedio causadas por un aumento en el tamaño de la muestra.

la vida útil y la calidad de las flores de *Phalaenopsis* híbridas” (por la confidencialidad de los resultados no se identifica el híbrido ni la empresa en la que se hizo el experimento).

Una de las limitantes más importantes para cumplir los planes de fertilización requeridos por las plantas de *Phalaenopsis* lo constituye la lluvia en los meses de agosto a noviembre puesto que si su frecuencia es muy alta, es necesario suspender la fertilización de las plantas, lo cual lógicamente afecta la producción de flores. Como solución al problema nutricional se planteó el uso de un sistema de riego para aplicar las soluciones fertilizantes y el uso de soluciones nutritivas con mayor concentración de nutrientes. Sin embargo, la aplicación de fertilizante por la vía foliar mediante sistemas de alta presión necesariamente moja las flores de la planta, por lo que surge el problema de que es posible que la concentración de sales de la solución nutritiva pueda disminuir la vida poscosecha de las flores al cambiar su composición interna como consecuencia de la absorción de la solución vía xilema o por efecto local por deshidratación sobre los pétalos de la flor al depositarse encima de su superficie gotas de la solución. Se hace entonces necesario probar el efecto de diferentes concentraciones de solución fertilizante sobre la calidad y vida útil de la flor. Así, la investigación que se efectuó buscaba alcanzar los siguientes objetivos:

- Determinar el efecto de la concentración de la solución nutritiva usada en la fertilización de las orquídeas sobre la calidad de las flores al momento de la cosecha.
- Determinar el efecto de la concentración de la solución nutritiva usada en la fertilización de las orquídeas, sobre la vida de florero de las flores después de una simulación de exportación.

Los objetivos enunciados llevaron al planteamiento, entre otras, de las siguientes hipótesis:

a.- El uso de solución nutritiva a diferentes concentraciones incrementa el número de flores en los escapos de *Phalaenopsis*. El planteamiento estadístico para la prueba de esta hipótesis es el siguiente:

Hipótesis nula: El número promedio de flores en los escapos de las plantas no sometidas a la fertilización (μ_0) es igual al número promedio de flores en los escapos de las plantas fertilizadas (μ_1).

Hipótesis alternativa: El número promedio de flores en los escapos de las plantas no sometidas a la fertilización es menor al número promedio de flores en los escapos de las plantas fertilizadas.

En símbolos:

$$H_0: \mu_0 = \mu_1$$

$$H_1: \mu_0 < \mu_1$$

Como se puede notar, la hipótesis que se desea rechazar se planteó en la nula y la que se desea probar en la alternativa.

b.- El uso de solución nutritiva no tiene efecto alguno sobre la vida útil de las flores en florero después de la exportación.

Hipótesis nula: El número promedio de días que tarda una flor en mostrar síntomas de deterioro es igual para las flores de plantas que no recibieron fertilización (μ_0) que para las flores de plantas que sí la recibieron (μ_1).

Hipótesis alternativa: El número promedio de días que tarda una flor en mostrar síntomas de deterioro

es mayor para las flores de plantas que no recibieron fertilización que para las flores de plantas que sí la recibieron.

$$H_0: \mu_0 = \mu_1$$

$$H_1: \mu_0 > \mu_1$$

Como se puede notar, la hipótesis que se desea probar se planteó en la nula, por lo tanto lo que se espera es que no se dé evidencia para rechazarla.

Se probaron como tratamientos cuatro concentraciones de la solución nutritiva usada por la empresa para la fertilización de las plantas: 0,5; 1, 2 y 3 gramos por litro.

El número de días transcurrido hasta la aparición del primer síntoma de deterioro resultó no ser significativo en el análisis de variancia.

Esto prueba que la concentración de la solución nutritiva que se aplicó como fertilización no tiene efecto alguno sobre el tiempo que la flor logra mantener su calidad de campo una vez cosechada. El promedio general de duración de los escapos fue de 3,77 días y un 75% de ellos tuvo una duración igual o mayor de tres días. La prueba utilizada mostró tener una potencia de 0,99 para detectar una diferencia de un día de un tratamiento a otro, por lo que sus resultados son bastante confiables.

La prueba de Dunnett mostró diferencias (con un nivel de significancia del 5%) entre el número promedio de flores del testigo y el de cada uno de los tratamientos evaluados. Los promedios del número de flores por escapo se muestran en el Cuadro 2.

Cuadro 2. Número promedio de flores por tratamiento.

Tratamiento	Número promedio de flores
Testigo	8,37
0,5 g/l	10,6
1 g/l	10,5
2 g/l	10,45
3 g/l	10,47

Se muestra una superioridad promedio de 2,14 flores más por escape de la solución nutritiva usada para la fertilización frente al testigo. Lo anterior es consecuencia del efecto del aporte de nutrientes que los tratamientos con solución nutritiva hicieron directamente sobre la flor.

Se concluye, por ende, que es posible fertilizar las *Phalaenopsis* utilizando un sistema de riego a presión sin disminuir la vida útil de las flores y que el uso de solución nutritiva aumenta la calidad (número de flores por escape) de las flores. Estas dos conclusiones tienen sustento científico. La primera está sustentada en una potencia de prueba del 0,99. Esto significa que si la prueba se repitiera un número muy grande de veces en las mismas condiciones, en 99 de cada 100 pruebas se detectarían una diferencia de un día si ésta existiera. Si se omite la mención a la potencia de la prueba quedaría la duda de si en realidad no hay una diferencia o si una diferencia real no se detectó por falta de potencia de prueba, que fue lo que se ilustró en el ejemplo hipotético. La segunda está sustentada en un nivel de significancia de 5%.

Lo que se ha presentado en este trabajo lleva a una última consideración con respecto a la fijación de los niveles de error para la prueba de una hipótesis. Cuando lo que se busca es rechazar la hipótesis nula, para una prueba estricta desde el punto de vista científico debe fijarse un nivel muy bajo para el error Tipo I (α). Si lo que se busca es demostrar una igualdad, en otras palabras no rechazar la hipótesis nula, fijar un alfa grande es más riguroso, puesto que, con los demás factores constantes, fijar un α grande hace más pequeño el error Tipo II (β) y por ende más alta la potencia de la prueba.

LITERATURA CITADA

- Daniel, WW. 2002. Bioestadística. Base para el análisis de las Ciencias de la Salud. Limusa Wiley. 755 p.
- Hernández, M; Herrera, F. 2005. Respuesta de los pastos *Panicum maximum*, *Brachiaria brizantha* y *B.decumbens* a herbicidas posemergentes con acción germicida. Revista de Agricultura Tropical 35:15-25.
- Levin, RI. 1988. Estadística para administradores. Prentice Hall. 866 p.
- Saborio, D; Mata, L; Arauz, LF. 1994. Tratamientos bactericidas poscosecha contra *Vibrio cholerae* en lechuga (*Lactuca sativa*). Agronomía Costarricense 18(1): 7-12.
- SAS Institute. 2000. JMP Statistical Discovery Software. Version 4. Statistics and Graphics Guide. 572 p.
- Sierra, R; Une, C; Ramírez, V; González, MI; Ramírez, JA; De Mascarel, A; Barahona, R; Salas-Aguilar, R; Páez, R; Avendaño, G; Ávalos, A; Broutet, N; Mégraud, F. 2006. Association of serum pepsinogen with atrophic body gastritis in Costa Rica. Clin.Exp.Med 6: 72-78.
- Senders, VL. 1958. Measurement and Statistics. Oxford University Press. 527 p.
- Wong, E; González, MI; Antillón, F; Glenn, E. 1997. Efecto de varios agentes, a diferentes niveles de pH, sobre la tasa de filtración de la piangua *Anadara tuberculosa* (Mollusca:Arcidae). Rev.Biol.Trop. 45 (4): 1453-145.