

Calf Notes.com

Calf Note #127 – ¿Qué significa "P menor que el punto 05"?

Introducción

Considere el siguiente escenario: acaba de terminar las tareas de la mañana en su rancho de terneros y se ha sentado en la silla frente a la computadora para ingresar algunos datos e imprimir un par de informes. El teléfono suena. "¡Buenos días! Esta es Sarah Bellim de Cheatham Feed Company. Me gustaría contarles acerca de nuestro último suplemento de reemplazo de leche para terneros, Calf Explosión. Nuestro último ensayo de investigación mostró una mejora estadísticamente significativa en las tasas de crecimiento de terneros alimentados con Calf explosion. ¿Puedo pasar mañana para contarte más sobre este emocionante producto nuevo?" Hmmm, piensas para ti mismo, estadísticamente significativo, ¿eh? Eso suena importante. Tal vez debería visitar a Sarah sobre este nuevo producto...

Pero, ¿qué significa realmente "estadísticamente significativo"? Y, ¿qué significa para usted como productor de terneros (o asesor)? Esta Calf Note está diseñada para abordar algunas de esas preguntas.

Hoy, se introducen nuevas tecnologías todos los días. Las mejoras en nutrición, información, reproducción, salud y manejo ocurren a un ritmo cada vez mayor. Cuando un productor incorpora estas nuevas tecnologías en la operación, se espera un cambio que mejore el rendimiento de los animales, reduzca el costo de producción o, de alguna manera, mejore la economía de la operación.

Pero, todos sabemos de nuevas tecnologías (productos y programas) que no funcionan. Puede haber muchas razones para la falta de un efecto, pero el hecho es que el producto no funciona en su operación (o tal vez no funciona en absoluto). Entonces, ¿cómo se separan los productos que son realmente efectivos de aquellos que son una pérdida de tiempo y dinero? El primer paso es comprender la ciencia detrás del producto. Y ahí es donde entra el término "estadísticamente significativo".

Variación: un ejemplo

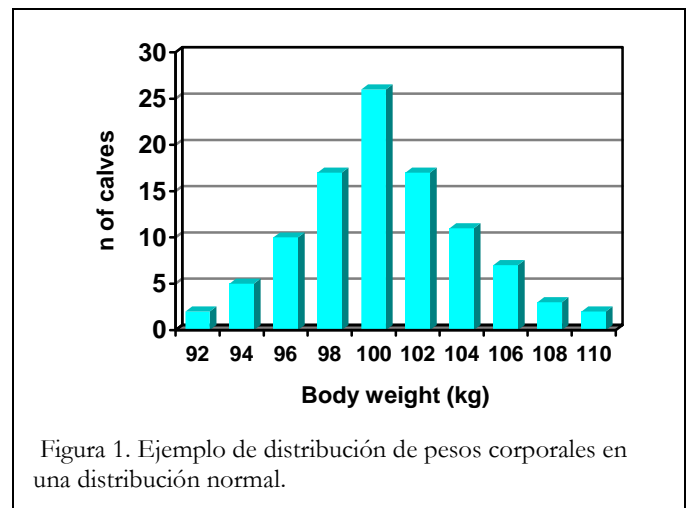
Veamos qué significa "estadísticamente significativo" a modo de ejemplo simple. Digamos que tienes dos grupos de 100 terneros. Al primer grupo alimentas una ración normal. Llamaremos a este grupo el grupo "Control". Para el segundo grupo, alimenta la ración normal más el aditivo de alimentación "X". Usted mide el peso corporal de los terneros al principio y nuevamente al final del estudio.

Veamos primero el grupo de control. Todos estos terneros fueron alimentados con la misma dieta, tratados de la misma manera y, en teoría, todos deberían pesar lo mismo al final del ensayo. Por supuesto, no todos los terneros pesan lo mismo. Algunas crías crecieron más rápido que otras debido a su composición genética normal. Otros terneros consumieron un poco más de sustituto de leche en algunos días que otros terneros. Aún otros terneros desarrollaron diarrea durante el ensayo,

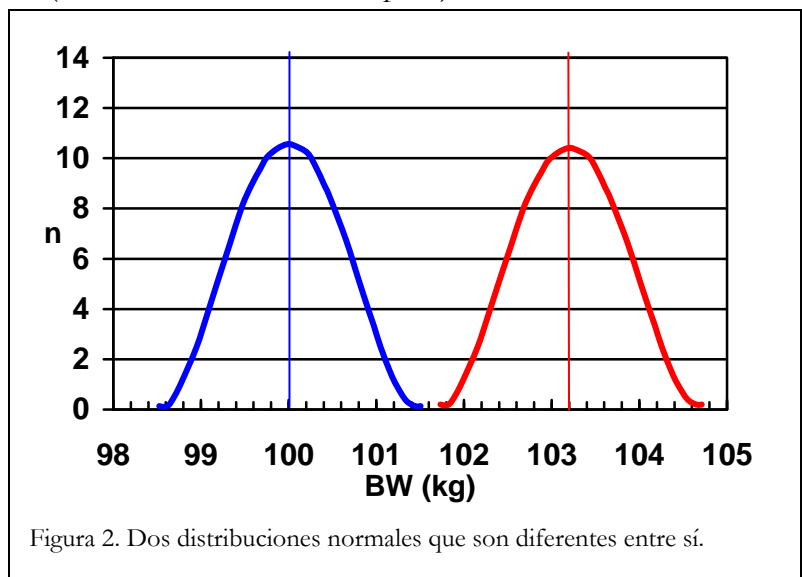
lo que desaceleró su crecimiento. Si tomamos el promedio del peso corporal final de todos los terneros de control, supongamos que el promedio es de 100 kg (220 lb). Pero, también hay una variación asociada con ese promedio, es decir, algunos terneros pesaron más que el promedio (también llamado la media) al final del estudio y algunos terneros pesaron menos que la media. Para este ejemplo, supongamos que todos los terneros pesaban entre 92 y 110 kg al final del estudio. Esta variación alrededor de la media se puede calcular de muchas maneras diferentes, pero un método común es calcular la desviación estándar de la media. Digamos que la desviación estándar de la media para el grupo de control fue de 4 kg. Si hacemos un gráfico del número de terneros en cada categoría de 2 kg (digamos, 92-93.9 kg, 94-95.9 kg), obtenemos un gráfico como la figura 1. Hay una variación normal y aleatoria en los pesos corporales y La distribución alrededor de la media (100 kg) se denomina "distribución normal". Este tipo de distribución también se denomina curva en forma de campana, ya que tiende a parecerse a una campana.

Ahora, hagamos lo mismo con los terneros alimentados con el nuevo aditivo. El peso corporal final promedio para este grupo de terneros fue de 103.2 kg con una desviación estándar de 5 kg.

Podemos usar los dos valores (la media y la desviación estándar) en cada grupo para determinar si la diferencia entre los dos grupos es realmente significativa, es decir, es estadísticamente significativa. Esto se hace usando una serie específica de procedimientos de estadísticas. Básicamente, estos procedimientos comparan los medios de cada tratamiento junto con la variabilidad en torno a estos medios para determinar si los dos medios son iguales (es decir, no hubo efecto del tratamiento "X") o si son diferentes.



Mire la Figura 2. Podemos ver rápidamente que no solo las medias son diferentes (las líneas rectas), sino que la variación alrededor de las medias (las líneas en forma de campana) también son bastante diferentes. Esto se debe a que no hay superposición entre las dos curvas en forma de campana. En la Figura 3, vemos que aunque las medias siguen siendo las mismas (100 y 103 kg), hay mucha más variación alrededor de las medias (es decir, las curvas de campana son mucho más anchas) y las distribuciones se superponen entre sí. En este caso, una prueba estadística concluiría que las medias NO son diferentes entre sí y el tratamiento "X" no hizo nada en nuestra prueba.



Dos tipos de errores

Cuando hacemos este tipo de pruebas, llegamos a una conclusión basada en las medias y las desviaciones estándar de cada grupo. Decidimos si los medios son los mismos o si son realmente diferentes. Y, por supuesto, existe la posibilidad de que nuestra decisión sea incorrecta. En este caso, hay dos formas posibles en que podríamos estar equivocados. El primer tipo de error (llamado error de tipo I) ocurre cuando concluimos que los medios son diferentes (hubo un efecto de tratamiento) cuando realmente no hubo ninguno. Esto se denomina falso positivo: un buen ejemplo es cuando un kit de prueba de embarazo le dice a una mujer que está embarazada cuando realmente no lo está.

El segundo tipo de error (llamado error tipo II) ocurre cuando concluimos que las medias son las mismas (sin efecto de tratamiento) cuando en realidad eran realmente diferentes. Este es un falso negativo: un ejemplo es cuando el kit de prueba de embarazo informa que una mujer no está embarazada cuando realmente lo está.

"Estadísticamente significativo"

Por convención, la mayoría de las pruebas estadísticas controlan el riesgo de falsos positivos al decir que queremos estar en lo correcto el 95% de las veces cuando concluimos que los medios de

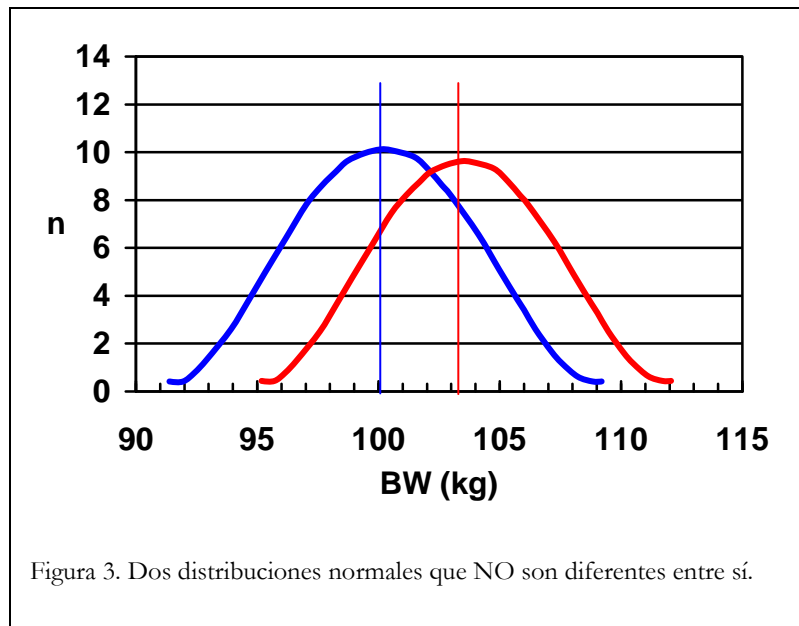
tratamiento son diferentes. Esto también significa que nos equivocaremos (es decir, concluimos que hubo una diferencia entre los tratamientos cuando no existía) solo el 5% del tiempo. Aquí es de donde proviene " $P < 0.05$ ". La probabilidad de llegar a una conclusión incorrecta es del 5% o menos.

Entonces, cuando alguien dice que las dos medias son diferentes en $P < 0.05$, significa que si el experimento se repitiera 100 veces en las mismas condiciones, entonces el 95% de las veces los resultados serían similares al primer experimento. Si bien los valores medios pueden no ser exactamente los mismos, las diferencias relativas deberían ser similares.

Es tradición que el término "estadísticamente significativo" o "significativo" signifique $P < 0.05$. Sin embargo, algunos investigadores pueden considerar estadísticamente significativo que $P < 0.10$ en lugar de $P < 0.05$. En general, es más fácil mostrar "importancia" cuando utiliza un valor P más alto. Es importante que comprenda lo que alguien quiere decir cuando dice "significativo". La pregunta importante es "si es significativo, ¿a qué nivel de probabilidad: 5%, 10% u otro nivel?"

¿Qué significa esto para usted?

La gente lanza el término "estadísticamente significativo" con gran abandono en estos días. Sin embargo, hay algunas implicaciones y precauciones importantes que usted como consumidor de la información debe tener en cuenta. Aquí hay algunos puntos a considerar al mirar los resultados de los ensayos de investigación:



1. El objetivo de la mayoría de los investigadores es encontrar una diferencia estadísticamente significativa si existe. Esto significa que es importante para él o ella reducir la variación en el experimento tanto como sea posible (por ejemplo, tratando de obtener gráficos como la Figura 2 en lugar de la Figura 3). Para lograr esto, los investigadores controlan tantos aspectos de la investigación como sea posible: animales, dietas, manejo, vivienda, medio ambiente, etc. Este control aumenta la capacidad de un investigador para ver la diferencia, pero ¿qué efecto tiene esto en usted? La variabilidad en su operación puede ser mayor porque no puede controlar todas las variables como un investigador. Esta es la razón más común por la que los productores prueban nuevos productos pero no pueden ver las diferencias prometidas por los defensores de la nueva tecnología.
2. ¿En qué condiciones se realizó la investigación? Los investigadores pueden controlar artificialmente las condiciones del experimento para aumentar la posibilidad de obtener una diferencia estadística. Puede ser posible ver una mejora en el crecimiento desde el nacimiento hasta el destete cuando se agrega un aditivo al sustituto de leche alimentado a 454 g/d sin alimentar el iniciador de ternera agregado; sin embargo, ¿cuántos productores en realidad no alimentan a sus terneros durante las primeras ocho semanas de vida? Busque un manejo artificial en la investigación, que debería hacer sonar las alarmas sobre la verdadera viabilidad de la tecnología.
3. Tenga cuidado con las poblaciones. A menudo, los investigadores de terneros trabajan con terneros (especialmente antes del destete) porque los terneros son más baratos de comprar y usar en la investigación. Por lo general, suponemos que los datos recopilados con terneros serán aplicables a terneros. Pero, ¿es esto siempre cierto? Tenga en cuenta si los investigadores también usan razas diferentes a la que usa en su operación.
4. ¡Cuidado con los pequeños estudios! Las reglas estadísticas dicen que si tienes un pequeño número de animales de experimentación, es más difícil declarar significación estadística. Sin embargo, un pequeño número de animales también puede significar que los animales son mucho más uniformes, lo que reduce la variabilidad. Esto puede hacer que los resultados de la investigación sean menos aplicables en el "mundo real". En general, me gusta ver un mínimo de 25-30 terneros por tratamiento en investigaciones con terneros jóvenes. Veo los resultados de los estudios que usan 10 terneros o menos con precaución.
5. ¿Cuál es la magnitud de la diferencia entre tratamientos en investigación? Muchos investigadores se entusiasman con la significación estadística pero pierden de vista los números reales. Podría ser posible ver una diferencia significativa de 250 gramos de aumento de peso corporal por día: ¿qué significa esto para un productor? Incluso en las mejores condiciones, nunca podrá medir un aumento de 250 gramos en la ganancia diaria en la granja.
6. ¿Cuál es el valor P que se usa?
7. ¿Qué se está midiendo? Muchos investigadores de terneros miden parámetros tales como aumento de peso corporal, aumento de altura, eficiencia de alimentación, etc. ¿Está midiendo estos parámetros en su operación? ¿Son importantes para usted? Si sus terneros pesan 5 kg más al destete, ¿qué significa esto para usted desde un punto de vista económico? Considere las medidas en cuestión y cómo el cambio puede afectar su operación.

8. ¿Cuál es la propuesta de valor? Si invierte \$ 1 por ternero por día en una nueva tecnología, ¿qué espera recuperar? ¿Cuál es el ROI (retorno de la inversión)? El beneficio económico dividido por el costo. Por ejemplo, usted invierte \$ 10,000 por año en una tecnología que reduce la pérdida de muerte de terneros destetados en un 5%, lo que le ahorra \$ 50,000 por año. El ROI es $50,000 / 10,000 = 5: 1$. La mayoría de los productores buscan un ROI mínimo de 3: 1

Los datos que provienen de estudios bien diseñados pueden ayudar a los productores a comprender el valor potencial de una nueva tecnología. Para comprender realmente si las diferencias entre las tecnologías "antiguas" y "nuevas" son reales, el análisis estadístico adecuado es esencial y lo ayudará a separar el "trigo" de la "paja". Sin embargo, hay muchas formas de "mentir con las estadísticas". Simplemente porque alguien le dice que la tecnología que está vendiendo resultó en una "diferencia significativa" que no significa que, bajo las condiciones de su operación, debe esperar ver la misma magnitud de diferencia si adopta la tecnología . La mejor de las suertes!

Escrito por Dr. Jim Quigley (30 de Septiembre de 2007)
© 2007 por Dr. Jim Quigley
Calf Notes.com (<http://www.calfnotes.com>)