

CONCEPTOS Y METODOS ESTADISTICOS INVOLUCRADOS EN LA EXPERIMENTACION DE SUELOS Y RESPUESTA VEGETAL

Walter E. Baethgen*

1. INTRODUCCION

La estadística es una disciplina relativamente reciente (del siglo XX), que comenzó a ser ampliamente utilizada en ciencias biológicas a partir de 1925 con las primeras publicaciones de Fisher. A partir de ese momento la estadística comenzó a formar parte de una gran cantidad de aspectos de la vida cotidiana. Para citar dos ejemplos típicos, por un lado es utilizada en todos los sistemas de control de calidad industrial, por otro lado es la herramienta básica para todas las encuestas de opinión y/o de preferencias.

La estadística y el análisis de datos son herramientas básicas para el agrónomo trabajando en investigación. Toda la experimentación llevada a cabo en los centros de investigación se realiza utilizando métodos estadísticos desde el comienzo (al definir y diseñar los tipos de experimentos), hasta el procesamiento de datos y análisis finales de los resultados.

Agrónomos extensionistas y técnicos asesores también requieren una base en estas áreas. Por un lado porque la fuente de información técnica fundamental son los resultados de investigación, cuyo análisis e interpretación requiere una base de conocimientos estadísticos. Por otro lado, es cada vez más frecuente la necesidad que los técnicos tienen de analizar grandes bases de datos provenientes de sus productores, cooperativas, empresas, etc., y también aquí conceptos de estadística aplicados constituyen una herramienta de gran utilidad.

En este marco, el principal objetivo de la presentación es refrescar conceptos teóricos

relacionados con elementos básicos de Estadística y técnicas de análisis de datos. Se intenta dar un sentido aplicado y no teórico, de tal forma que puedan ser utilizados en la actividad cotidiana de técnicos desempeñando actividades de extensión y asesoramiento.

2. DISTRIBUCIONES Y CARACTERIZACION DE LAS MISMAS

En primer lugar se tratan nociones básicas de distribuciones. Se enfatiza en la distribución normal, por ser la que más frecuentemente se utiliza en la Agronomía. Se describe dicha distribución, y la distribución de frecuencias acumuladas como una forma sencilla de visualizar probabilidades de ocurrencia.

Las dos formas más sencillas de caracterizar una distribución, es con medidas de tendencia central y de dispersión. Entre los parámetros más típicos para caracterizar una distribución se presentan la media, la mediana, los cuartiles, deciles, percentiles, etc.

Por otro lado, al caracterizar la variabilidad de una distribución, se discute cuáles serían las características ideales de dicho parámetro. Por ejemplo: (a) el parámetro debería ser pequeño si los valores están cerca de la media, y grande si están lejos de la misma; (b) debería ser independiente del tamaño de los valores (unidades); (c) debería permitir interpretar su valor, es decir debería dar una idea de cuán común o poco común es una observación en una población definida. En este sentido, el mejor parámetro para definir variabilidad de una población es la varianza y su raíz cuadrada el desvío estándar.

*Ph.D., M.Sc., Research and Development Division, International Fertilizer Development Center, EEUU.

Es decir que conociendo la "forma" en que se distribuyen los valores de una población o muestra, y sabiendo los valores de tendencia central y variabilidad, podemos caracterizar dicha población o muestra. En otras palabras, podemos tener una buena idea de cuán común es un determinado valor, cuál es el rango de valores más comunes, etc. La estadística incluye una serie de herramientas que utilizan estas medidas de tendencia central, y de variabilidad de una o varias muestras, para inferir conclusiones de poblaciones mayores.

3. EXPERIMENTOS, ERROR EXPERIMENTAL Y MODELO LINEAL ADITIVO

Se introduce el concepto de experimento como pruebas planeadas para descubrir, verificar o rechazar resultados. Los tipos de experimentos más comunes son: (a) preliminares, donde se prueban muchas alternativas generalmente sin repeticiones; (b) críticos, los más comunes en centros de investigación, en donde se incluyen algunas alternativas, y en donde se utilizan repeticiones; y (c) demostrativos en los que se incluyen pocas alternativas (generalmente las más promisorias de experimentos críticos), y que normalmente se utilizan con fines de extensión y/o de validación de resultados de investigación a nivel de productor.

Se vuelve entonces necesario introducir el concepto estadístico de error (o error experimental) como la causa o conjunto de causas que hacen que al repetir un mismo tratamiento el resultado no sea exactamente el mismo. Se enumeran los mecanismos de control del error experimental, y se dan ejemplos de como funcionan: tamaño de la parcela, medición de variables concomitantes, y diseño experimental.

Se estudia a continuación uno de los conceptos estadísticos más útiles e importantes: el Modelo Lineal Aditivo. Con ejemplos prácticos se intenta definir el modelo lineal aditivo, y se le utiliza como introducción a los diseños experimentales y a sus Análisis de Varianza. El Análisis de Varianza es una herramienta que, considerando el modelo lineal aditivo, particiona la variabilidad total en componen-

tes de variación con origen conocido. Por ejemplo, el modelo lineal aditivo puede indicar que el valor de una observación cualquiera es el fruto de un componente de la media de la población de donde fue extraída la observación (es decir el valor más esperado para dicha observación), más un componente debido al tratamiento a que fue sujeta esa observación (por ej.: una dosis de fertilizante, un cultivar de trigo, etc.), más un componente de variación que no puede ser explicado (que puede ser debida a errores en la toma de muestra, a la variabilidad "natural" de la población de donde proviene, etc.).

4. ANALISIS DE VARIANZA, DISEÑO EXPERIMENTAL Y SEPARACION DE MEDIAS

El análisis de varianza cuantifica cada uno de los mencionados componentes de la variación total, y permite interpretar resultados en cuanto a los efectos que los tratamientos tuvieron sobre la población en que se está trabajando. Si el o los tratamientos que estamos probando no tienen ningún efecto sobre la población, entonces la variabilidad de la población luego de aplicados los tratamientos debería ser similar a la variabilidad de la población original. Si en cambio alguno o todos los tratamientos afectan a la población, luego de aplicados estos tratamientos van a originar una variabilidad mayor que la de la población original. El análisis de varianza compara la variabilidad de la población que se debe a la aplicación de los tratamientos con la variabilidad original mediante la prueba "F".

Se describen también en esta sección los diseños de Parcelas al Azar, Bloques al Azar, y Parcelas Divididas, con sus propiedades y Análisis de Varianza correspondientes. La prueba "F" se estudia en detalle interpretando sus resultados y limitaciones, la selección de un determinado nivel de significancia, etc.

Las pruebas "F" resultantes de los análisis de varianza sólo nos permiten concluir que (a) no hay diferencia entre los tratamientos (si F es no significativa), o (b) que por o menos dos tratamientos son diferentes. Sin embargo no nos dice nada en cuanto a cuál o cuáles tratamientos son diferentes, ni cuán impor-

tante son las diferencias entre tratamientos. Para estos fines es necesario utilizar técnicas de separación de medias. En esta sección se describen las características más importantes de las técnicas más comúnmente utilizadas en investigación agropecuaria: la mínima diferencia significativa de Fisher, y las pruebas de Duncan y Tuckey. Se estudian también las limitaciones de las mencionadas técnicas para separar medias de tratamientos estructurados (estructuras de grupo, gradiente, o factoriales). Se introduce entonces el concepto de contrastes. Finalmente en esta sección se describen los experimentos factoriales: es decir experimentos en los que se estudia el efecto de dos o más factores simultáneamente. Se introduce el concepto de interacción que permite evaluar el efecto de un factor para diferentes niveles de otro(s) factor(es). Se demuestra con ejemplos prácticos las limitaciones de las técnicas de separación de medias convencionales en el análisis de estos experimentos, y de la importancia de analizar cuidadosamente los resultados cuando la interacción es importante.

5. MODELOS DE REGRESION SIMPLE, CUADRATICA Y MULTIPLE

La siguiente sección está destinada a la introducción de modelos de regresión y su análisis. Se comienza por el modelo de regresión simple, y se pone especial énfasis en la interpretación de los resultados (prueba F, coeficiente de correlación, coeficiente de determinación, etc.). En esta sección se explica lo que es cada componente del análisis de regresión desde el punto de vista práctico. Se trata luego el modelo de regresión cuadrática como uno de los más comúnmente incluidos en la investigación en Fertilidad de Suelos. Nuevamente se recalca en la interpretación de los coeficientes de regresión en forma individual. En especial se enfatizan los posibles errores que pueden cometerse al intentar caracterizar una regresión cuadrática considerando solamente el coeficiente lineal.

Finalmente, para el caso de regresiones cuadráticas se estudia también la capacidad de analizar los resultados de investigación

desde el punto de vista económico. Es decir, si la regresión que estamos estudiando es de la forma:

$$Y = a + bX + cX^2$$

El valor de X que maximiza Y es:

$$X_{\max} = b / 2c$$

Y el valor de X que optimiza el resultado económico de Y es:

$$X_{\text{opt}} = (b - RP) / 2c$$

donde RP es la relación de precios de X:Y (o sea la cantidad de unidades de Y necesarias para comprar una unidad de X).

Aquí es donde el coeficiente lineal (b) es importante: cuanto mayor sea la diferencia entre el valor de "b" y de "RP" menos va a cambiar el óptimo si la relación de precios cambia. En cambio cuando el valor de "b" y de la relación de precios son similares, el resultado del óptimo económico va a cambiar mucho si la relación de precios se modifica.

Finalmente, se introduce el concepto de regresión múltiple, y se incluyen ejemplos de como simplificar su análisis e interpretación.

BIBLIOGRAFIA RECOMENDADA

- DRAPER, N.R.; SMITH, H.** 1966. Applied regression analysis. New York, Wiley.
- SNEDECOR, G.W.; COCHRAN, G.** 1980. Statistical methods. Ames, The Iowa State University Press.
- STEEL, R.G.D.; TORRIE, J.A.** 1980. Principles and procedures of statistics. New York, McGraw-Hill.