

APLICACIONES DE LA ESTADÍSTICA EN PRODUCCIÓN Y SANIDAD ANIMAL

Vet. Alberto Morillo Alujas, Doctor Ing. Agr. Daniel Villalba Mata, Ing. Téc. Agrícola Jaume Feixa Cercós.
Tests & Trials, S.L., Altorricón, Huesca.

www.produccion-animal.com.ar

INTRODUCCIÓN

El papel de la estadística en la investigación está fuera de toda duda. Pero, quizás, no le demos una mayor relevancia en cualquiera de nuestras tareas cotidianas. Seguramente, hemos usado expresiones tales como: cuando hay que escoger a un macho reproductor, "escojo al mejor en cuanto a productividad", o esta otra: "mis animales son mejores que los tuyos", sin pensar entre cuántos animales y cuánto mejor que los otros y qué sentido tiene el término "mejor". En estos dos casos, no estamos haciendo otra cosa que aplicar el pensamiento estadístico.

Creemos que es indispensable definir la estadística antes que adentrarnos en ella. Una definición clara y sencilla podría ser el método científico de operar con los datos y interpretarlos. De esta forma podremos recopilar datos, analizarlos, agruparlos y presentarlos en cuadros o gráficos para su posterior interpretación.

TÉRMINOS HABITUALES EN ESTADÍSTICA

Para irnos familiarizando con la Estadística, es importante definir una serie de conceptos básicos. El primero de ellos es la población, que podría ser definida como el conjunto de individuos que tienen una característica en común (por ejemplo, todos los lechones destetados el mismo día), pudiendo ser finitas o infinitas. El siguiente es el término dato, como medida o valor que puede ser observado y contado (por ejemplo, el peso de un lechón cualquiera es de 4.6 gramos). A continuación definiremos muestra, siendo una parte o subconjunto de la población (por ejemplo, seleccionaremos cinco lechones de cada cerda); se toman muestras cuando es difícil o costosa la observación de toda la población. Por último tenemos el término variable, que sería una característica que puede adoptar diferentes valores en los diferentes individuos (por ejemplo, si antes hemos escogido como dato el peso de un lechón, el peso sería una variable).

TIPOS DE VARIABLES

Las variables estadísticas las podemos clasificar según sean:

Variables discretas. Son aquellas que toman valores aislados, es decir, sólo puede tomar ciertos valores enteros y exactos. Como ejemplos se pueden citar el número de animales enfermos de una explotación (si son 10 animales enfermos, no pueden ser 13), o el número de lechones de una cerda.

Variables continuas. Son aquellas que pueden tomar infinitos valores en un intervalo dado por muy próximo que sea el intervalo. El peso de los lechones, la ganancia media diaria de los cerdos, serían ejemplos de variables estadísticas continuas.

RECOPIACIÓN DE LA INFORMACIÓN

Cuando se definió el término población se hizo referencia al carácter finito e infinito de esta. Luego, al explicar el término muestra, nos referimos a ella como la imposibilidad de observar el conjunto de la población (a causa de recursos limitados, escasez de la población, destrucción de la población). Por esta razón, en Estadística se trabaja con las muestras, debido a que pueden tener cualquier tamaño pero teniendo en cuenta que las conclusiones y los resultados harán referencia a la población. Trabajar con muestras está supeditado a que éstas sean representativas de la población y que sean aleatorias. Al decir que las muestras sean aleatorias, estamos diciendo que cualquier integrante de la muestra tiene tantas probabilidades de ser incluido en ella como cualquier otro. Sería preciso hacer unas breves e intuitivas pinceladas entorno el tamaño de muestra. La situación ideal sería poder medir la población entera, así tendríamos la información en su totalidad. A medida que vayamos disminuyendo la población (muestrear), estaremos de acuerdo que perderemos información. Nos interesa que el tamaño de muestra sea lo mayor posible, teniendo en cuenta las restricciones antes expuestas, de esta forma la muestra será más aleatoria y representativa de la población. Por ejemplo, al tirar diez monedas, el resultado más probable sería obtener cinco caras y cinco cruces; pero la probabilidad de obtener más de cinco caras (o cruces) es mayor que si el número de lanzamientos es de 1000.

ORDENACIÓN DE LOS DATOS. DISTRIBUCIÓN DE FRECUENCIAS

Una vez obtenidos los datos, nos daremos cuenta que son numerosos y que hará falta organizarlos y clasificarlos de acuerdo a algún método estadístico, puesto que una variable estadística (por ejemplo, el peso de los lechones) puede tomar distintos valores y cada uno de éstos puede aparecer repetido más de una vez (frecuencia de los datos). Este método es la distribución de frecuencias, que nos permite el estudio del comportamiento de los datos de la población..

Por otro lado, precisaremos de algún esquema que nos permita tratar los datos de una forma manejable. Uno de estos métodos es la representación gráfica (por ejemplo, histogramas, diagramas de barras de la distribución de frecuencias). Otro, sería un esquema numérico que describa breve y exactamente la distribución de frecuencias. Las cantidades que proporciona este esquema se conocen como estadísticos descriptivos.

REPRESENTACIÓN GRÁFICA

VARIABLES DISCRETAS

Las representaciones más adecuadas para este tipo de datos son las tablas de frecuencias, diagramas de barras i gráficos de sectores.

VARIABLES CONTINUAS

Una representación sencilla de estos datos son las tablas de frecuencia o los polígonos de frecuencia. Otra forma de representación es el diagrama acumulativo. Otras herramientas gráficas son los pictogramas y por último los histogramas.

ESTADÍSTICOS DESCRIPTIVOS

DE LOCALIZACIÓN

A veces es conveniente reducir la información a un solo valor o a un número pequeño de valores para comparar muestras o poblaciones. De alguna forma estamos centralizando la información, por esta razón se les conoce como medidas de tendencia central.

Media aritmética

Este estadístico es muy importante. Puede adoptar el nombre de promedio. Se calcula sumando todos los datos individuales y dividiéndolo por el número de datos de la muestra.

$$\text{Ej. } X = \{1,5,12,9,6,5,10\}$$

$$\text{Media} = (1+5+12+9+6+5+10) / 7 = 48 / 7 = 6,857$$

Mediana

La consideraremos el valor central de una distribución de frecuencias. De esta forma la mediana nos divide la distribución en dos mitades.

$$\text{Ej. } X = \{1,5,12,9,6,5,10\}$$

$$\text{Mediana} = 9$$

Moda

Es el valor de la variable que tiene máxima frecuencia. No tiene por que ser única.

$$\text{Ej. } X = \{1,5,12,9,6,5,10\}$$

$$\text{Moda} = 5$$

DE DISPERSIÓN

Puede ocurrir que dos distribuciones de frecuencias puedan tener la misma media aritmética. Estos estadísticos nos dan una alternativa para caracterizar las distribuciones.

Rango

Representa la diferencia entre el mayor valor y el menor de las observaciones de una muestra. Nos da una ínfima idea de la dispersión de las observaciones.

Desviación típica

La podríamos definir como la diferencia entre una observación individual y la media aritmética. Nos da una idea de la variabilidad de los datos de nuestra muestra. Valores pequeños de la desviación nos dicen que es poco probable que la variable que estamos estudiando se desvíe de la media aritmética, es decir, la dispersión es baja.

Coeficiente de variación

La desviación típica sólo la podemos usar para hacernos una idea de la variabilidad de una única muestra o población. No tendría sentido comparar las desviaciones típicas de dos muestras o poblaciones diferentes. En cambio el coeficiente de variación nos permite comparar la variación existente entre dos muestras o poblaciones con media aritmética diferente. La podemos considerar como el cociente entre la desviación típica y la media aritmética.

De asimetría

Para poder explicar estos estadísticos creemos necesario primero mencionar cual es la distribución de frecuencias más importante, que no es otra que la distribución normal. Muchas de las variables que encontramos en la naturaleza siguen esta distribución. La gráfica de una distribución normal es una curva acampanada que se extiende a lo largo del eje X, tanto en sentido positivo como negativo.

Asimetría

Diremos que una distribución de frecuencias es simétrica cuando los valores de la media aritmética, mediana y moda coinciden. La asimetría (no coincidencia de las medidas de tendencia central) puede darse hacia la derecha (positiva) o hacia la izquierda (negativa). Cuando la distribución presenta asimetría se dice que está sesgada.

Curtosi

Las curvas de distribución, comparadas con las curvas de distribución normal, pueden presentar diferentes grados de apuntamiento o de altura de la cima de la curva. En algunos casos una distribución puede presentar sus valores concentrados cerca de la media, presentando un pico elevado. En otros casos, la distribución puede ser relativamente plana. Una medida de curtosi sirve para apreciar el grado en que una curva de distribución de frecuencias es más alta o más achatada que la curva normal de distribución.

ESTADÍSTICA MEDIANTE EXCEL

El objetivo de este trabajo no ha sido sustituir un curso de estadística, sino hacer unas breves pinceladas de las partes más importantes de la estadística. Por esta razón se ha obviado la explicación de conceptos estadísticos mediante fórmulas farragosas y complicadas. Se ha creído conveniente sustituir estas fórmulas por el cálculo mediante Excel, programa informático al abasto de cualquier usuario y relativamente sencillo de usar.

Ingresar en el programa

Desde Windows 95, para entrar en el programa Excel, desplácese con el mouse hasta el botón inicio (parte inferior izquierda de la pantalla) y presiónelo. Luego seleccione la opción de menú programas y presione Microsoft Excel.

Distribución y funcionamiento del programa Excel

Está formado por celdas, cada una de las cuales está identificada por una letra (parte superior de la hoja) y un número (lado izquierdo de la hoja). Las celdas pueden contener valores, títulos o fórmulas. En la parte superior de la hoja, encontraremos una barra de menú, empezando por el menú archivo, seguido del menú insertar, etc. En el menú insertar será donde se hallarán las funciones. Dentro de funciones se desplegará un cuadro donde encontraremos las funciones estadísticas. Una vez presionada la función estadística deseada, se desplegará un cuadro que nos permite seleccionar las celdas que queremos operar (deberemos apartar dicho cuadro arrastrándolo con el mouse y presionando la primera celda, sin soltar el mouse, arrastrar hasta la última celda que queramos seleccionar). Para realizar cualquier operación en Excel, es obligado situarse en una celda.

Funciones estadísticas

MAX

Determina el máximo del rango de celdas que seleccionemos.

MIN

Determina el mínimo del rango de celdas que seleccionemos.

MEDIANA

Halla la mediana del rango de celdas.

MODA

Muestra el valor modal del rango de celdas que hayamos seleccionado.

PROMEDIO

Calcula la media aritmética del rango seleccionado.

DESVESTP

Calcula la desviación típica de los valores contenidos en las celdas seleccionadas.

COEFICIENTE. ASIMETRIA

Calcula el coeficiente de asimetría para el rango de celdas que seleccionamos.

DISTR.NORM(x, media, desv-estándar, Acum)

Calcula el valor de la distribución acumulada o el valor de la función de densidad de probabilidad de una variable aleatoria con distribución normal con media y desviación estándar especificadas. El cuarto argumento (Acum) puede ser verdadero o falso; si usamos verdadero se obtiene el valor para la distribución acumulativa, si se utiliza Falso se obtiene el valor de la función de densidad de probabilidad. A efectos prácticos usaremos Acum= FALSO.

Gráfica de la distribución normal estándar

La gráfica de la distribución normal se puede realizar tomando valores en el eje X comprendida entre -2 y 2; por ejemplo, podemos tomar -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, colocándolos en una columna de un hoja Excel y en la columna siguiente escribiendo la fórmula para cada X (por ejemplo, para X=-2, usaríamos la formula DISTR.NORM(-2,0,1,falso). A continuación, con el mouse presionaremos la opción insertar del menú principal, y seleccionaremos la opción gráfico. El siguiente paso sería seleccionar el tipo de gráfico (en este caso puede ser líneas) y hacemos clic en el botón rotulado siguiente. En el campo correspondiente a rango de datos escribiremos el rango de celdas donde se encuentran los datos (por ejemplo, si se encuentran entre la celda A1 y A9, escribiremos A1:A9) y clicar siguiente. El siguiente paso sería terminar, a menos que queramos adornar el gráficos con leyendas, ejes, etc.

Ejemplo

Se conocen los pesos al destete de los lechones de dos granjas diferentes (granja A y granja B).

Peso al destete de la granja A (gramos)	Peso al destete de la granja B (gramos)
7,5	5,1
6,4	3,9
7,1	4,3
2,9	7,03
4,6	4,6
9,4	5,7
7,5	6,4
6,8	4,9
3,9	3,8
4,1	5,9
4,6	6,2
5,2	3,8
6,8	5,7
5,2	4,7
6,4	5,3

Se pide:

Reconocer las variables y decir de qué tipo de variables se tratan.

Calcular los estadísticos de localización, de dispersión y de asimetría.

Comparar las dos granjas en función de los estadísticos calculados.

Solución

Las variables que intervienen en este ejemplo son el peso de los lechones al destete en la granja A y el peso de los lechones al destete en la granja B. Se trata de variables continuas debido a que pueden adoptar cualquier valor en un intervalo dado.

El cálculo de los estadísticos se ha realizado con el programa Excel, siguiendo los pasos descritos.

	Granja A	Granja B
Media aritmética	5,89	5,16
Mediana	6,40	5,10
Moda	7,50	5,70
Máximo	9,40	7,03
Mínimo	2,90	3,80
Desviación típica	1,66	0,96
Coefficiente asimetría	0,15	0,21
Coefficiente variación	0,28	0,19

El coeficiente de variación ha sido calculado como el cociente entre la desviación típica y la media aritmética.

Primeramente nos debemos fijar en las medias de las dos granjas. Podemos comprobar como las dos medias son muy parecidas, siendo algo superior la media de la granja A. A continuación, fijémosnos en la desviación típica. Que la desviación típica sea una u otra, en principio, no nos da mucha información. Pero observemos que la desviación de la granja A es superior a la de B; esto nos da una primera idea que en la granja A hay más variabilidad en los datos obtenidos. Esta idea tendrá que ser corroborada mediante el coeficiente de variación. La primera idea que tuvimos se ha cumplido. El coeficiente de variación de la granja A es mayor que el de la granja B; una posible interpretación es que los datos de la granja A están más dispersos, más alejados del valor de la media aritmética, mientras que los datos de la granja B se encuentran más concentrados alrededor de la media, es decir, son valores más uniformes.