

Aplicación didáctica de los conceptos de correlación y regresión lineal a un caso en Ciencias Veterinarias

D. C. Sánchez¹, J. Fernández Madero^{1,2}

Resumen

La enseñanza de la estadística en el primer año de Ciencias Veterinarias presenta desafíos constantes a docentes y alumnos.

Se muestra, en este trabajo una experiencia didáctica para la enseñanza del procedimiento de análisis de correlación y regresión lineal. La enseñanza se apoya en datos producto de investigaciones de docentes de la cátedra, aunque el alumno simula su participación desde la adquisición de los pares aleatorios de mediciones. Se aplican las técnicas estadísticas adecuadas y secuenciadas, con la permanente referencia práctica de los resultados en términos de un problema de calidad de carne que se plantea como motivador.

La experiencia resultó de gran aceptación entre los estudiantes que además integraron estos contenidos de bioestadística con otros aplicados a las ciencias veterinarias.

Palabras clave: didáctica - correlación - regresión - aplicaciones - veterinaria

Introducción

La estadística no es una forma de hacer sino una forma de pensar, que ayuda a la solución de problemas en las ciencias, pero también en la vida cotidiana. La enseñanza de la estadística se debe iniciar con problemas reales donde los estudiantes puedan desarrollar sus ideas, trabajando las diferentes etapas que conlleva la resolución de un problema real (planificar la solución, recolectar y analizar los datos, comprobar las hipótesis iniciales y tomar una decisión en consecuencia) (1).

El conocimiento de la metodología estadística es un arma imprescindible para la obtención, análisis e interpretación de los datos que proceden de observaciones sistemáticas o de experimentaciones proyectadas, específicamente para conocer los efectos de uno o varios factores que intervienen en los fenómenos en estudio (2).

Una de las grandes dificultades en la enseñanza de la estadística es la heterogeneidad que se presenta en el alumnado, ya que los estudiantes no tienen las mismas bases matemáticas, ni el mismo hábito de razonar, sumándose a esto el

1. Cátedra de Bioestadística. csanchez@ucasal.net

2. Cátedra de Bromatología

nivel de prevención con el que los estudiantes llegan a la materia por desconocimiento o porque constatan la necesidad de contar con algunas habilidades matemáticas básicas en las que no se sienten afianzados. Por esta razón, el docente debe buscar constantemente nuevas estrategias didácticas, que motiven y comprometan al estudiante a ser partícipe de la construcción de su propio aprendizaje.

Moreno Echavarría (4) propone un modelo constructivista de enseñanza del tema, basado en la participación del alumno desde el momento de la toma de las mediciones de variables antropométricas para la obtención de la serie de datos a analizar por los métodos de correlación y regresión.

El análisis de correlación lineal estudia el grado y sentido de la asociación lineal que hay entre un conjunto de variables y, a diferencia del análisis de regresión no persigue estimar explícitamente el modelo funcional, sino la fuerza de la asociación.

El análisis de regresión es conocido como una técnica estadística que permite modelar la relación entre variables, es decir, el estudio de asociación cuantitativa entre un número de variables. Su objetivo principal, es explorar la relación existente entre las variables para obtener información de una de ellas a través del conocimiento de los valores de la otra.

En referencia al abordaje del tema de Correlación y Regresión Lineal en los libros de texto recomendados como bibliografía de Cátedra, Walpole et al. (5) inicia el capítulo de regresión lineal con la definición de las variables y la explicación de los términos a utilizar tales como la ecuación de regresión y el coeficiente de regresión; luego se introduce el tema de regresión lineal simple a través de un ejemplo en donde se halla la relación existente entre la reducción de

sólidos y la demanda química de oxígeno en 50 observaciones en efluentes de curtiembre, las cuales son tabuladas y posteriormente graficadas usando el diagrama de dispersión. También se habla de los mínimos cuadrados como la técnica matemática para hallar los valores llamados por Walpole coeficientes de regresión.

En el texto de Di Rienzo et al. (3), el tema se introduce con la mención de algunos ejemplos prácticos de análisis de las relaciones funcionales entre variables como la cantidad de suplemento administrado a animales y su incremento de peso o la dosis de un insecticida y la mortalidad de los insectos tratados y problematiza con interrogantes como: «¿existe alguna relación entre estas variables? ¿Qué forma tiene esa relación? ¿Es posible predecir los valores de una variable en función de otra?, etc.

En todos los casos la aproximación se inicia con la representación gráfica de las variables a través de un diagrama de dispersión, donde resulta sencillo identificar intuitivamente el tipo de relación directa o indirecta, aunque a la mayoría de los estudiantes, les resulta dificultoso construir este diagrama desde la asignación de las escalas de medición de las variables en ambos ejes.

La adquisición de los conceptos de correlación y regresión lineal para su aplicación en el campo de la veterinaria son esenciales en la formación del estudiante, para que posteriormente, en su formación de postgrado adquirir nuevos conocimientos con esta base, tales como su relación con otras técnicas estadísticas como el Análisis de la Varianza, el análisis de matrices de correlación múltiple, los modelos de regresión lineal múltiple y la detección de autocorrelación en matrices multivariadas.

Frecuentemente, los métodos de regresión se aplican al análisis de datos que provienen de estudios que no fueron diseñados, este es el caso

del estudio de fenómenos no controlados o de registros históricos; pero también son muy útiles en experimentos diseñados, en los que se factoriza a distintos niveles los tratamientos.

El objetivo del trabajo es presentar un abordaje constructivista para la enseñanza de los conceptos de regresión y correlación lineal, en base a datos reales obtenidos en investigaciones desarrolladas en la Facultad de Ciencias Agrarias y Veterinarias de la Universidad Católica de Salta.

Materiales y métodos

Se identificaron series de datos que servirían para ilustrar los conceptos de correlación y regresión lineal.

En un estudio sobre satisfacción del consumidor de carne se midió la dureza en muestras de carne provenientes de individuos de raza Brangus sometidos a condiciones experimentales preestablecidas que se relatan a los estudiantes con auxilio de diapositivas.

La terneza (antagónica de la dureza) es la cualidad de la carne de dejarse cortar y masticar

(con mayor o menor facilidad) antes de la deglución, estando directamente ligada a la resistencia mecánica del producto consumible. La terneza de la carne está dada por múltiples factores intrínsecos y extrínsecos.

Esta variable de respuesta, la dureza, fue medida por dos métodos: por el aparato de corte en cizalla Warner-Bratzler de la Facultad de Veterinaria de la Universidad de Buenos Aires y la evaluación sensorial a través del panel de expertos de la misma institución que determinaron el número de masticaciones antes de la deglución. De acuerdo a la bibliografía, ambos métodos pueden correlacionarse o no.

Se orientó un juego de roles donde un grupo es el panelista de degustadores y otro el personal que opera el aparato de corte en cizalla para simular cómo se obtiene cada par de determinaciones por muestra, mientras una docente procede a la carga de los datos en el software estadístico InfoStat v.2012.

El procedimiento de análisis se presenta en la Figura 1.

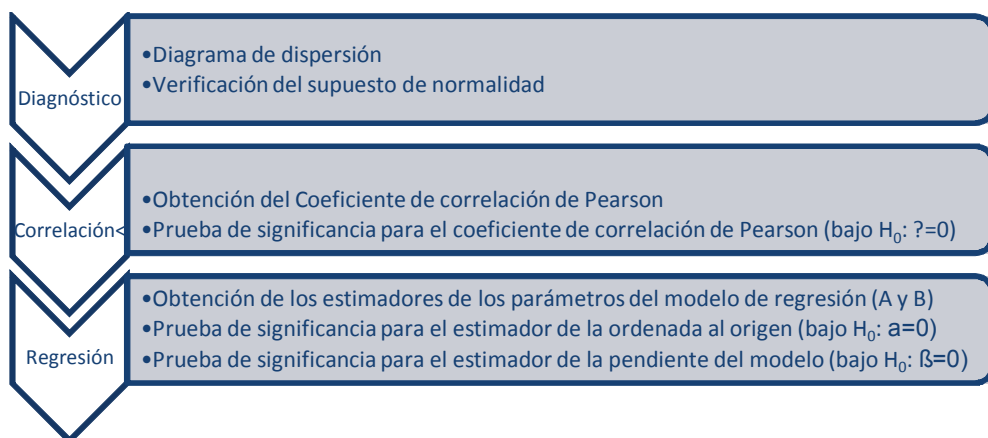
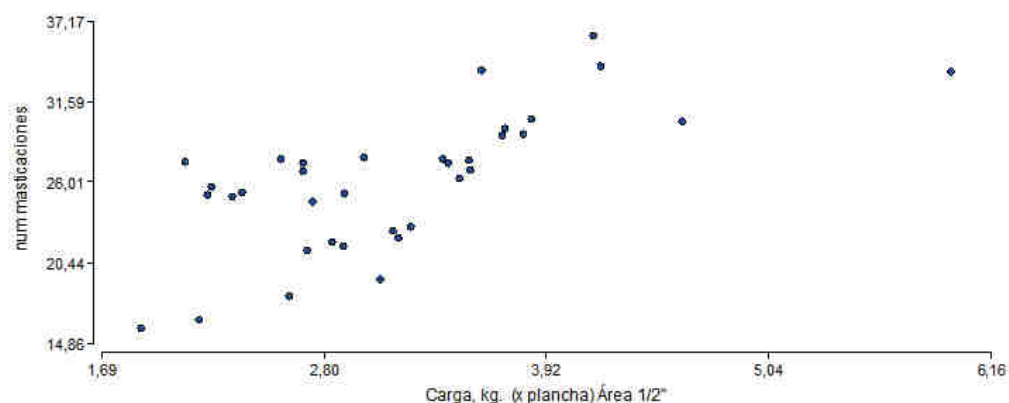


Figura 1. Procedimiento de análisis de los pares de datos de Dureza WB y número de masticaciones

Resultados

Se obtuvo una representación gráfica de las variables a través de un diagrama de dispersión, donde pudo observarse el tipo de relación existente entre las variables en análisis. Del análisis

gráfico se deduce colectivamente que la relación entre las variables es de naturaleza directa: a medida que aumenta la dureza medida por el instrumento de corte, también crece el número de masticaciones antes de la deglución



Se realizó el análisis de la correlación utilizando el coeficiente de correlación de Pearson observándose una existencia de correlación lineal significativa entre la Dureza WB (obtenida por el instrumento de corte) y el número de masticaciones y se alcanzó un $r=0,69$ que resultó significativo para el test de significancia ($H_0: \rho=0$) con $p<0,0001$, de modo que la asociación entre la dureza medida por instrumental y el número de masticaciones resultó altamente significativa. A la hora de tomar decisiones se repasaron conceptos previos como los pasos del test de hipótesis para esta prueba.

Para la aplicación del coeficiente de correlación de Pearson, se hizo una previa verificación del supuesto de normalidad de distribución de los residuos, a través del test de normalidad de Shapiro Wilks modificado.

Para afianzar el concepto de que correlación significativa no implica causalidad, se explicaron

los fundamentos de los dos tipos de mediciones, haciendo hincapié en el efecto de medir en el instrumental con un corte en cizalla y el número de masticaciones obtenido en el panel de expertos.

En una etapa posterior, confirmada la asociación significativa entre las variables se plantea un modelo análisis de regresión lineal, donde la variable independiente es la dureza medida por instrumental y la variable dependiente es el número de masticaciones.

Para encontrar la relación funcional entre las variables, se propuso un modelo de regresión lineal simple del tipo

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Donde:

Y_i : observación de la variable dependiente (número de masticaciones) de la i -ésima mues-

tra, donde $i=1,2,\dots,n$. En el caso analizado $n=35$.

X_i : es el i -ésimo valor de la variable independiente (valor de dureza en la muestra i).

a : parámetro que representa la ordenada al origen de la recta poblacional (valor de Y cuando $X=0$)

β : parámetro que representa la pendiente de la recta (tasa de cambio del número de masticaciones frente al cambio unitario de la dureza)

ε_i : variación aleatoria (o no explicada por el modelo) asociada a la i -ésima observación de Y bajo el nivel X_i .

Este modelo incluye sólo una variable independiente (regresión lineal simple) y establece que la esperanza de la variable dependiente cambia con tasa constante, según crece o decrece el valor de la variable independiente.

Se obtuvieron los coeficientes de regresión para obtener la función que explica la relación entre las variables.

$$\text{Número de Masticaciones} = 26,05 + 4,03 \text{ Dureza WB}$$

A los fines predictivos esta función tiene una limitación: sólo puede aplicarse en el rango de variación de dureza medida en el instrumento de corte en cizalla WB entre 2,35 y 5,95.

Posteriormente se aplicó un test de significancia para los estimadores de los parámetros del modelo (a y β), que resultaron altamente significativos ($p < 0,0001$).

Discusión

La construcción de los conocimientos de correlación y regresión lineal a través de un ejemplo concreto, desde la obtención del archivo de pares de datos, hasta la verificación de la significancia del modelo, resultó sumamente motivador para los alumnos y docentes de la Cátedra.

En el futuro, debería intentarse llevar la experiencia incluso a la medición de las variables para relacionar con otros conceptos como precisión, fuentes de error, etc. y para hacer más completa para participación del alumno.

Conclusiones

Las bases de datos generadas en las tareas de investigación de los docentes y donde hay participación de alumnos avanzados, son fuentes de excelentes ejemplos para enseñar estadística aplicada.

Volcar los resultados de investigaciones de docentes en los ejercicios y desafíos al alumno en el aula, no sólo asegura la transferencia y enriquecimiento de la propuesta pedagógica con conocimientos recientemente generados, sino que permiten al alumno vislumbrar el sentido de aprender y aplicar estadística para los médicos veterinarios.

Referencias

1. Batanero C. 2001. *Didáctica de la Estadística. Grupo de Educación Estadística*. Departamento de Didáctica de la Matemática. Universidad de Granada.
2. Box, George E. P.; Hunter, William G.; Hunter, Stuart J. 1999. *Estadística para investigadores*. Reverté, S.A., México.

3. Di Rienzo J., Casanoves F., Gonzalez L., Tablada M., Díaz M., Robledo C. Balzarini M. 2008. *Estadística para las Ciencias Agropecuarias*. Editorial Brujas. Córdoba.
4. Moreno Chavarría R.B. 2012. *Propuesta didáctica para la enseñanza de la estadística en los modelos de regresión lineal simple bajo un enfoque constructivista*. Tesis de Maestría en la Enseñanza de las Ciencias Exactas y Naturales. Universidad Nacional de Colombia. Medellín.
5. Walpole R.E., Myers R.H., Myers S.L. (1999) *Probabilidad y Estadística para ingenieros*. Prentice Hall Hispanoamericana, S.A. 6º Edición. México.

Agradecimientos

Se agradece la invaluable colaboración de Luis Marcelo Torino Solá, estudiante de Ingeniería en Producción Agropecuaria en la Pontificia Universidad Católica Argentina, quien cedió gentilmente parte de los datos para la realización de este trabajo.