Rev.MVZ Córdoba 19(2):4099-4108, 2014. ISSN: 0122-0268

ORIGINAL

Validation of models with constant bias: an applied approach

Validación de modelos con sesgo constante: un enfoque aplicado

Salvador Medina-Peralta,[1] M.Sc, Luis Vargas-Villamil,[2]* Ph.D, Jorge Navarro A,[3] Ph.D, Leonel Avendaño R,[4] Ph.D, Luis Colorado M,[1] M.Sc, Enrique Arjona-Suarez,[5] Ph.D, German Mendoza-Martinez,[6] Ph.D.

[1]Universidad Autónoma de Yucatán, Facultad de Matemáticas, Apartado Postal 172 Cordemex, C.P. 97119, Mérida, Yucatán, México. [2]Colegio de Postgraduados, Campus Tabasco, Periférico Carlos A. Molina Km. 3.5, Apartado Postal 24, C.P. 86500, Cárdenas, Tabasco, México. [3]Universidad Autónoma de Yucatán, Facultad de Medicina Veterinaria y Zootecnia, Apartado Postal 4-116 Itzimná, C.P. 97100, Mérida, Yucatán, México. [4]Universidad Autónoma de Baja California, Instituto de Ciencias Agrícolas, 17 Boulevard Delta s/n, Ejido Nuevo León, C.P. 21705, Baja California, México. [5]Colegio de Postgraduados, Programa en Estadística, Km 36.5 Carretera México-Texcoco, C.P. 56230, Montecillo, México. [6]Universidad Autónoma Metropolitana, Xochimilco, Departamento de Producción Agrícola y Animal, Medicina Veterinaria y Zootecnia, Edificio 34, 3er piso, Calzada del Hueso 1100 Col Villa Quietud 04960, México D.F., México. Correspondencia: luis@avanzavet.com

Received: January 2013; Accepted:November 2013.

ABSTRACT

Objective. This paper presents extensions to the statistical validation method based on the procedure of Freese when a model shows constant bias (CB) in its predictions and illustrate the method with data from a new mechanistic model that predict weight gain in cattle. **Materials and methods.** The extensions were the hypothesis tests and maximum anticipated error for the alternative approach, and the confidence interval for a quantile of the distribution of errors. **Results.** The model evaluated showed CB, once the CB is removed and with a confidence level of 95%, the magnitude of the error does not exceed 0.575 kg. Therefore, the validated model can be used to predict the daily weight gain of cattle, although it will require an adjustment in its structure based on the presence of CB to increase the accuracy of its forecasts. **Conclusions.** The confidence interval for the 1-α quantile of the distribution of errors after correcting the constant bias, allows determining the top limit for the magnitude of the error of prediction and use it to evaluate the evolution of the model in the forecasting of the system. The confidence interval approach to validate a model is more informative than the hypothesis tests for the same purpose.

Key words: Statistics, error, confidence interval, models (*Source: MeSH*)

RESUMEN

Objetivo. Presentar extensiones al método estadístico para validar modelos basado en el procedimiento de Freese cuando el modelo presenta sesgo constante (SC) en sus predicciones e ilustrar el método con datos provenientes de un modelo mecanístico inédito para la predicción de ganancia de peso de bovinos. **Materiales y métodos.** Las extensiones fueron la prueba de hipótesis y error máximo anticipado para el planteamiento alternativo y el intervalo de confianza para un cuantil de la distribución de los errores. **Resultados.** El modelo evaluado presentó SC, una vez eliminado y con un nivel de confianza del 95%, la magnitud del error no sobrepasa 0.575 kg. Por lo que el modelo validado puede usarse para predecir la ganancia de peso diaria de bovinos, aunque requerirá un ajuste en su estructura con base a la presencia de SC para incrementar la exactitud en sus pronósticos. **Conclusiones.** El intervalo de confianza para el cuantil 1-α de la distribución de los errores una vez que se corrige el sesgo constante, permite determinar una cota superior para la magnitud del error de predicción y usarla para evaluar la evolución del modelo en predicción del sistema. El enfoque de intervalos de confianza para validar un modelo es más informativo que las pruebas de hipótesis para el mismo propósito.

Palabras clave: Estadística, error, intervalo de confianza, modelos (*Fuente: MeSH*).

INTRODUCTION

The validation of a model is defined as the comparison of the forecasts of the model with the values observed in the actual system to determine whether the model is suitable for the established purpose (1, 2). In the mathematical modeling process, the validation stage with observed data different to those used for obtaining the parameters of the model, plays a fundamental role in the models that will be applied and where forecasts will be used instead of the measurements of the actual system, which may be too costly or difficult to obtain. Barrales et al (3) indicate that in modeling of systems, an essential stage, which poses both conceptual and practical difficulties, is the validation of the models.

Different approaches and techniques have been presented in the literature to validate models. For Mayer and Butler (4), validation techniques can be grouped into four main categories: subjective evaluation (involves a number of experts in the field of interest), visual techniques (comparative graphs), measures of deviation (based on the differences between the values observed and simulated) and statistical tests. In turn, Tedeschi (5) conducted a review of various techniques to evaluate mathematical models designed for predictive purposes: linear regression analysis, adjusted errors analysis, concordance correlation coefficient, several evaluation measurements, mean squared prediction error, non-parametric analysis and comparison of the distribution of the data observed and predicted. Medina-Peralta, Vargas-Villamil (6), indicate that some measures of deviation for validating models contradict with the graphical methods when the model is biased in its forecasts; they recommend using

INTRODUCCIÓN

La validación de un modelo se define como la comparación de las predicciones del modelo con los valores observados del sistema real para determinar si el modelo es adecuado para el propósito establecido (1, 2). En el proceso de modelación matemática, la etapa de validación con datos observados diferentes a los utilizados en la obtención de los parámetros del modelo, juega un papel fundamental para modelos que serán aplicados y donde las predicciones serán empleadas en lugar de las mediciones del sistema real, las cuales pueden ser demasiado costosas o difíciles de obtener. Barrales et al (3) señalan que en la modelación de sistemas, una etapa esencial y que presenta dificultades tanto conceptuales como prácticas, es la validación de los modelos.

En la literatura se han expuesto diferentes enfoques y técnicas para validar modelos. Para Mayer y Butler (4), las técnicas de validación se pueden agrupar en cuatro categorías principales: evaluación subjetiva (involucra a un número de expertos en el campo de interés), técnicas visuales (gráficas comparativas), medidas de desviación (basadas en las diferencias entre valores observados y simulados) y pruebas estadísticas. Por su parte, Tedeschi (5) realizó una revisión de varias técnicas para evaluar modelos matemáticos diseñados para propósitos predictivos: análisis de regresión lineal, análisis de los errores ajustados, coeficiente de correlación de concordancia, diversas medidas para evaluación, cuadrado medio del error de predicción, análisis no paramétricos y la comparación de la distribución de los datos observados y predichos. Medina-Peralta, Vargas-Villamil (6), señalan que algunas medidas de desviación para validar modelos se contradicen

a graphical methods and measures of deviation to validate models.

Freese (7) presented a statistical procedure for validating models without bias (SS) and with constant bias (SC) or proportional bias (SP) in forecasts, which essentially involves determining whether the accuracy and precision of a model or estimation technique meets the requirements of the modeler or user of the model. Rennie and Wiant (8), based on the procedure of Freese (7), proposed the use of an maximum anticipated error or critical error to determine if a model is accurate and precise. Reynolds (9) reviewed the assumptions, deduced the procedure of Freese and extended the results to the case of SS models: a) an alternative approach to the Chi-square test for the accuracy test, b) another critical error based on the alternative approach, and c) a confidence interval (CI) for the quantile 1-α of the distribution of errors. Barrales et al (3) extended the procedure of Freese (7) and the modification suggested by Rennie and Wiant (8) in two ways: a) explicit expression of the procedures of the statistical test under the original approach, for models with SC or SP when the maximum permissible error is expressed as a percentage of the actual values, and b) indication of how to obtain the critical error under the original approach when the model presents SC or SP.

This paper presents the extensions to the statistical method to validate the models proposed by Freese (7), Rennie and Wiant (8), Reynolds (9) and Barrales et al (3) when the model shows SC in its predictions. The method is illustrated with data from a new mechanistic model for the prediction of weight gain in cattle.

MATERIALS AND METHODS

Basic concepts. Let's suppose that there are "n" pairs to compare (Y_i, Z_i) i=1,2,…,n, where for the pair i, Y_i is the observed value, Z_i the corresponding predicted value and $D_i=Y_i-Z_i$ the difference between them. On Freese's approach (7) for the determination of the accuracy and precision required, the e values (maximum error allowed in deviations $|y_i-z_i|=|d_i|$) and α (1-α represents the level of accuracy required) specified by the modeler or user of the model, must satisfy that D conforms to a normal distribution with mean zero and $P(|D|\leq e)\geq(1-\alpha)$ for the model to be acceptable and be considered as sufficiently reliable for the prediction of the system. Thus, a model is accurate or unbiased (SS) when it is true that D conforms to a normal distribution with zero mean.

con los métodos gráficos cuando el modelo presenta sesgo en sus pronósticos; recomiendan utilizar conjuntamente medidas de desviación y métodos gráficos para validar modelos.

Freese (7) presentó un procedimiento estadístico para validar modelos sin sesgo (SS) y con sesgo constante (SC) o sesgo proporcional (SP) en sus pronósticos, el cual consiste fundamentalmente en determinar si la exactitud y precisión de un modelo o técnica de estimación satisface los requerimientos del modelador o usuario del modelo. Rennie y Wiant (8) basados en el procedimiento de Freese (7), proponen el empleo de un error máximo anticipado o error crítico para determinar si el modelo es exacto y preciso. Reynolds (9) revisó los supuestos, dedujo el procedimiento de Freese y extendió resultados al caso de modelos SS: a) un planteamiento alternativo a la prueba ji-cuadrada para la prueba de precisión, b) otro error crítico basado en el planteamiento alternativo, y c) un intervalo de confianza (IC) para el cuantil 1-α de la distribución de los errores. Barrales et al (3) extendieron el procedimiento de Freese (7) y la modificación sugerida por Rennie y Wiant (8) de dos formas: a) expresión explícita de los procedimientos de la prueba estadística bajo el planteamiento original, para modelos con SC o SP cuando el máximo error permitido se expresa como un porcentaje de los valores reales, y b) indicación de cómo obtener el error crítico bajo el planteamiento original cuando el modelo presenta SC o SP.

En este trabajo se presentan extensiones al método estadístico para validar modelos planteados por Freese (7), Rennie y Wiant (8), Reynolds (9) y Barrales et al (3) cuando el modelo presenta SC en sus predicciones. Se ilustra el método con datos provenientes de un modelo mecanístico inédito para la predicción de ganancia de peso de bovinos.

MATERIALES Y MÉTODOS

Conceptos básicos. Suponga que se tienen "n" pares para comparar (Y_i, Z_i) i=1,2,…,n donde para el i-ésimo par, Y_i es el valor observado, Z_i el valor predicho correspondiente y $D_i=Y_i-Z_i$ la diferencia entre ellos. En el planteamiento de Freese (7) para la determinación de la exactitud y precisión requerida, se necesita que los valores e (máximo error admitido de las desviaciones $|y_i-z_i|=|d_i|$) y α (1-α representa el nivel de exactitud requerida) especificados por el modelador o usuario del modelo satisfagan respectivamente que D se ajusta a una distribución normal con media

Determination of constant bias. According to the graphical method used by Barrales et al (3) and Medina-Peralta, Vargas-Villamil (6), the presence of SC will be identified in the predictions of the model. The SC is recognized by a value of the average of differences ($\bar{d}$) away from zero and for the graph of the dots (z_i, $d_i=y_i-z_i$) forming a horizontal band centered on $\bar{d}$ with a systematic distribution to positive or negative (points above and below the line d= $\bar{d}$) (6, 10). A statistical test for determining if the average of the differences is different from zero would contribute to determine the SC more objectively.

Validation of models with constant bias. When D has a normal distribution with variance σ_D^2 and mean μ_D [$D\sim N(\mu_D,\sigma_D^2)$] and the model is not accurate for having SC, the precision required $P(|D|\leq e)\geq 1-\alpha$ translates into $\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ once the SC is eliminated by correcting $W=D-\bar{D}$ (10). The correction makes the model to accurate and only one statistical test would be precision. The next step is to test the hypothesis with the original approach $H_0:\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ vs $H_1:\sigma_D^2 > e^2/\chi^2_{1,1-\alpha}$ (7, 10) or the alternative approach $H_0^{PA}:\sigma_D^2 > e^2/\chi^2_{1,1-\alpha}$ vs $H_1^{PA}:\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ (10). Reject H_0 o H_0^{PA} with a level of significance α' if:

$$V^{SC}=\frac{\chi^2_{1,1-\alpha}}{e^2}\left[\sum_{i=1}^{n}(d_i-\bar{d})^2\right] > \chi^2_{n-1,1-\alpha'} \text{ o } V^{SC} \leq \chi^2_{n-1,\alpha'} \quad (1)$$

where in general $\chi^2_{k,\gamma}$ represents the quantile γ of the chi-square distribution with "k" degrees of freedom, i.e., $P(X_k^2 \leq \chi^2_{k,\gamma})=\gamma$. Therefore, if $H_0$is not rejected or H_0^{PA} is rejected, then the model is considered acceptable to forecast under the original and alternative approach, respectively.

Another approach to evaluate precision is through confidence intervals. Thus, when $D\sim N(\mu_D,\sigma_D^2)$ and the model shows SC, the critical errors corresponding to the original and the alternative approach are:

$$e^{*}_{SC}=\left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,1-\alpha'}}\right)^{1/2} \quad (2)$$

$$e^{**}_{SC}=\left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,\alpha'}}\right)^{1/2} \quad (3)$$

These were obtained by isolating "e" from the rejection regions in equation (1) for H_0 and H_0^{PA} (10); the relationship existing between hypothesis tests and confidence intervals can be found in Casella and Berger (11), which allows constructing one from the other. Therefore, if the cero y $P(|D|\leq e)\geq(1-\alpha)$, para que el modelo sea aceptable y éste sea considerado suficientemente confiable para predicción del sistema. Así, un modelo es exacto o sin sesgo (SS) cuando se cumple que D se ajusta a una distribución normal con media cero.

Determinación de sesgo constante. Conforme al método gráfico, empleado por Barrales et al (3) y Medina-Peralta, Vargas-Villamil (6), se identificará la presencia de SC en las predicciones del modelo. El SC es reconocido por un valor del promedio de las diferencias ($\bar{d}$) alejado del cero y que el gráfico de los puntos (z_i, $d_i=y_i-z_i$) forme una banda horizontal centrada alrededor de $\bar{d}$ con una distribución sistemática a ser positivos o negativos (puntos arriba y debajo de la recta d= $\bar{d}$) (6,10). Una prueba estadística para determinar si la media de las diferencias es diferente de cero contribuiría a determinar el SC de manera más objetiva.

Validación de modelos con sesgo constante. Cuando D tiene una distribución normal con media μ_D y varianza σ_D^2 [$D\sim N(\mu_D,\sigma_D^2)$] y el modelo no es exacto por presentar SC, la precisión requerida $P(|D|\leq e)\geq 1-\alpha$ se traduce en $\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ una vez que es eliminado el SC mediante la corrección $W=D-\bar{D}$ (10). La corrección lleva a que el modelo sea exacto y sólo faltaría una prueba estadística para la precisión. El siguiente paso es probar las hipótesis con el planteamiento original $H_0:\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ vs $H_1:\sigma_D^2 > e^2/\chi^2_{1,1-\alpha}$ (7, 10) o el planteamiento alternativo $H_0^{PA}:\sigma_D^2 > e^2/\chi^2_{1,1-\alpha}$ vs $H_1^{PA}:\sigma_D^2 \leq e^2/\chi^2_{1,1-\alpha}$ (10). Rechace H_0 o H_0^{PA} con un nivel de significación α' si:

$$V^{SC}=\frac{\chi^2_{1,1-\alpha}}{e^2}\left[\sum_{i=1}^{n}(d_i-\bar{d})^2\right] > \chi^2_{n-1,1-\alpha'} \text{ o } V^{SC} \leq \chi^2_{n-1,\alpha'} \quad (1)$$

donde en general $\chi^2_{k,\gamma}$ representa el cuantil γ de la distribución ji-cuadrada con "k" grados de libertad, es decir, $P(X_k^2 \leq \chi^2_{k,\gamma})=\gamma$. Por tanto, si H_0 no se rechaza o H_0^{PA} se rechaza entonces el modelo es considerado aceptable para predicción bajo el planteamiento original y el alternativo respectivamente.

Otro enfoque para evaluar la precisión es por medio de intervalos de confianza. Así, cuando $D\sim N(\mu_D,\sigma_D^2)$ y el modelo presenta SC los errores críticos correspondientes al planteamiento original y al alternativo son:

$$e^{*}_{SC}=\left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,1-\alpha'}}\right)^{1/2} \quad (2)$$

modeler or user of the model specifies a value for "e" such that $e \geq e^{*}_{SC}$ o $e \geq e^{**}_{SC}$ then the model is considered acceptable for predicting the system under the original or the alternative approach, respectively.

An estimate with an CI of 100(1-α')% for $\varepsilon_{SC}=(\sigma_D^2 \chi^2_{1,1-\alpha})^{1/2}$ (10), the quantile 1-α of the distribution of errors (e) once the SC is corrected through $w_i=d_i-\bar{d}$, is given by:

$$\left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,1-\frac{\alpha'}{2}}}\right)^{1/2} < \varepsilon_{SC} < \left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,\frac{\alpha'}{2}}}\right)^{1/2} \quad (4)$$

$$e^{*}_{ISC} < \varepsilon_{SC} < e^{**}_{ISC}$$

The CI estimated (e^{*}_{ISC}, e^{**}_{ISC}) means that there is a confidence of 100(1-α')% that the point of the distribution $|D-\bar{D}|$ which must be under 100(1-α)% of the absolute errors is located somewhere in such interval. It allows determining with a certain probability an upper bound for the magnitude of the prediction error and use it to evaluate the evolution of the model in the prediction of the system.

RESULTS

To illustrate the application of the methodology, data from a new dynamic mechanistic model called *Wakax POS* were used for the prediction of the average weight gain (AWG) per day of bovine animals in a tropical area of Mexico (Table 1). This model was developed by Ph.D. Luis Vargas Villamil of Colegio de Postgraduados, Campus Tabasco, in postdoctoral studies at the Department of Animal Science of the University of California, Davis (UCD). The model *Wakax POS* describes the biological relationships (digestion, bacterial growth, fermentation and absorption) during the nutrition of cattle fed with sugarcane (CZ) and predicts the average weight gain (AWG) per day in grazing cattle supplemented with CZ, broken corn and/or molasses in a tropical area of Mexico. It consists of 119 states variables that describe the system composed of five sub-models: concentrates, grass, sugarcane, molasses and animal growth. The input variables of the model are: a) live weight; b) consumption of dry matter of corn, molasses, and grass; c) soluble fraction of grass and CZ; d) biodegradable fraction of grass and CZ; and e) ratio of grass and CZ degradation (10).

The goodness of fit tests of Cramer-von Mises ($W^2=2.172$, $p<0.003$) and Anderson-Darling ($A^2=10.733$, $p<0.003$) reject ($p<0.05$) the hypothesis $D\sim N(0,\sigma_D^2)$ with unspecified variance, hence the model is not accurate.

$$e^{**}_{SC} = \left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,\alpha'}}\right)^{1/2} \quad (3)$$

Estos se obtuvieron al despejar "e" de las regiones de rechazo en la ecuación (1) para H_0 y H_0^{PA} (10); en Casella y Berger (11) puede consultarse la relación existente entre pruebas de hipótesis e intervalos de confianza, la cual permite construir uno a partir del otro. Por lo tanto, si el modelador o usuario del modelo especifica un valor de "e" tal que $e \geq e^{*}_{SC}$ o $e \geq e^{**}_{SC}$ entonces el modelo es considerado aceptable para predecir el sistema bajo el planteamiento original o el alternativo respectivamente.

Una estimación por IC del 100(1-α')% para $\varepsilon_{SC}=(\sigma_D^2 \chi^2_{1,1-\alpha})^{1/2}$ (10), el cuantil 1-α de la distribución de los errores (e) una vez que se corrige el SC mediante $w_i=d_i-\bar{d}$, está dado por:

$$\left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,1-\frac{\alpha'}{2}}}\right)^{1/2} < \varepsilon_{SC} < \left(\frac{\chi^2_{1,1-\alpha}\sum_{i=1}^{n}(d_i-\bar{d})^2}{\chi^2_{n-1,\frac{\alpha'}{2}}}\right)^{1/2} \quad (4)$$

$$e^{*}_{ISC} < \varepsilon_{SC} < e^{**}_{ISC}$$

El IC estimado (e^{*}_{ISC}, e^{**}_{ISC}) significa que se tiene confianza en un 100(1-α')% que el punto de la distribución $|D-\bar{D}|$ que tiene debajo el 100(1-α)% de los errores absolutos está localizado en alguna parte de dicho intervalo. Permite determinar con cierta probabilidad una cota superior para la magnitud del error de predicción y usarla para evaluar la evolución del modelo en predicción del sistema.

RESULTADOS

Para ilustrar la aplicación de la metodología se utilizaron datos provenientes de un modelo dinámico mecanístico inédito llamado *Wakax POS* para la predicción de la ganancia de peso promedio (GPP) por día de bovinos en una zona tropical de México (Tabla 1). Este modelo fue desarrollado por el Dr. Luis Vargas Villamil del Colegio de Postgraduados, Campus Tabasco, en una estancia posdoctoral en el Departamento de Ciencia Animal de la Universidad de California, Davis (UCD). El modelo *Wakax POS* describe las relaciones biológicas (digestión, crecimiento bacteriano, fermentación y absorción) durante la nutrición de bovinos alimentados con caña de azúcar (CZ) y predice la ganancia de peso promedio (GPP) por día de bovinos en pastoreo suplementado con CZ, maíz quebrado y/o melaza en una zona tropical de México. Consta de 119 variables de estado que describen el sistema

Table 1. Weight gain averages (kg) observed and predicted in 34 experiments.

Observed (y_i)	Modeled (z_i)	Difference ($d_i=y_i-z_i$)	Correction for SC ($w_i=d_i-0.233$)
0.366	0	0.366	0.133
0.398	0	0.398	0.165
0.4930	0.390	0.103	-0.13
0.630	0.430	0.200	-0.033
0.430	0.130	0.300	0.067
0.400	0.280	0.120	-0.113
0.490	0.230	0.260	0.027
0.600	0.290	0.310	0.077
0	0.280	-0.280	-0.513
0.430	0.090	0.340	0.107
0.700	0.320	0.380	0.147
0.690	0.330	0.360	0.127
0.500	0.020	0.480	0.247
0.680	0	0.680	0.447
0.620	0	0.620	0.387
0.520	0.260	0.260	0.027
0.640	0	0.640	0.407
0.450	0	0.450	0.217
0.430	0.330	0.100	-0.133
0.760	0.390	0.370	0.137
0.036	0	0.036	-0.197
0.019	0	0.019	-0.214
0.412	0	0.412	0.179
0.052	0	0.052	-0.181
0.414	0	0.414	0.181
0.050	0	0.050	-0.183
0.292	0.350	-0.058	-0.291
0.054	0	0.054	-0.179
0.308	0.390	-0.082	-0.315
0.054	0	0.054	-0.179
0.037	0	0.037	-0.196
0.051	0	0.051	-0.182
0.062	0	0.062	-0.171
0.364	0	0.364	0.131

The next step is to test $D\sim N(\mu_D,\sigma_D^2)$ with unspecified mean and variance and identify the type of bias. The hypothesis $D\sim N(\mu_D,\sigma_D^2)$ is not rejected ($p>0.05$) (Cramer-von Mises: $W^2=0.118$, $p=0.065$; Anderson-Darling: $A^2=0.655$, $p=0.084$; Shapiro-Wilks: $W=0.961$, $p=0.254$; Kolmogorov-Smirnov: $KS=0.135$, $p=0.115$). In addition, the distribution of the points in figure 1 shows SC, because: (i) the mean of the differences (μ_D) is different from zero ($t=6.092$, g.l.=33, $p<0.001$), and (ii) the points (z_i, d_i) practically form a horizontal band centered on the line $\bar{d}=0.233$.

The last step is to apply the correction for SC for the model to be exact and test $W=D-\bar{D}\sim N(0,\sigma_W^2)$ with unspecified variance.

The goodness of fit tests of Cramer-von Mises ($W^2=0.122$, $p>0.25$) and Anderson-Darling ($A^2=0.675$, $p>0.25$) do not reject ($p>0.05$) the hypothesis $W\sim N(0,\sigma_W^2)$.

In this example both approaches were applied, that of the hypothesis tests and that of the confidence intervals, when the model shows SC in its forecasts.

compuesto por cinco submodelos: concentrado, pasto, caña de azúcar, melaza y crecimiento animal. Las variables de entrada del modelo son: a) peso vivo; b) consumo de materia seca de maíz, melaza y pasto; c) fracción soluble de pasto y CZ; d) fracción degradable de pasto y CZ; y e) razón de degradación de pasto y CZ (10).

Las pruebas de bondad de ajuste de Cramér-von Mises ($W^2=2.172$, $p<0.003$) y Anderson-Darling ($A^2=10.733$, $p<0.003$) rechazan ($p<0.05$) la hipótesis $D\sim N(0,\sigma_D^2)$ con varianza no especificada, por lo que el modelo no es exacto.

El siguiente paso es probar si $D\sim N(\mu_D,\sigma_D^2)$ con media y varianza no especificadas e identificar el tipo de sesgo. La hipótesis $D\sim N(\mu_D,\sigma_D^2)$ no se rechaza ($p>0.05$) (Cramér-von Mises: $W^2=0.118$, $p=0.065$; Anderson-Darling: $A^2=0.655$, $p=0.084$; Shapiro-Wilks: $W=0.961$, $p=0.254$; Kolmogorov-Smirnov: $KS=0.135$, $p=0.115$). Además, la distribución de los puntos en la figura 1 muestra SC, ya que: (i) la media de las diferencias (μ_D) es diferente de cero ($t=6.092$, g.l.=33, $p<0.001$), y (ii) los puntos (z_i, d_i) forman prácticamente una banda horizontal centrada alrededor de la recta $\bar{d}=0.233$.

El siguiente paso es aplicar la corrección por SC para que el modelo sea exacto y probar si $W=D-\bar{D}\sim N(0,\sigma_W^2)$ con varianza no especificada. Las pruebas de bondad de ajuste de Cramér-von Mises ($W^2=0.122$, $p>0.25$) y Anderson-Darling ($A^2=0.675$, $p>0.25$) no rechazan ($p>0.05$) la hipótesis $W\sim N(0,\sigma_W^2)$.

En este ejemplo se aplicaron ambos enfoques, el de pruebas de hipótesis y el de intervalos de confianza, cuando el modelo presenta SC en sus predicciones. Adicionalmente, se calculó el IC del 95% para ε_{SC}, el cuantil $1-\alpha$ de la distribución de los errores (e) una vez que se

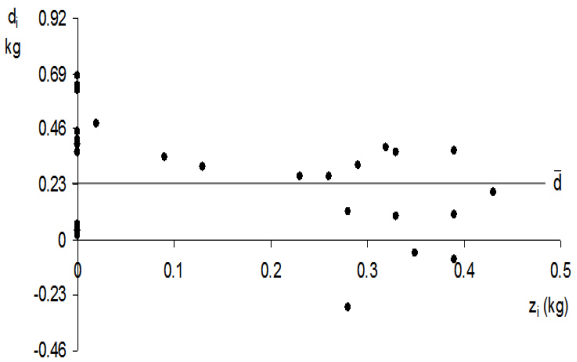

Figure 1. Relationship between the bias (d_i) and the values simulated (z_i) for average weight gain, where $\bar{d}$ is the average of the differences (d_i).

Additionally, a 95% CI was calculated for ε_{SC}, quantile 1-α of the distribution of errors (e) once the SC is corrected by $w_i=d_i-\bar{d}$. With $\alpha=\alpha'=0.05$ y e=0.5 kg as the maximum error admitted by the modeler or the user of the model, we have that: a) $P(|D|\leq 0.5kg)\geq 0.95$, b) $\sigma_D^2 \leq (e^2/\chi^2_{1,1-\alpha})=0.065$, c) $V^{SC}=25.220$ and a level of significance p=0.832 with the original approach (PO) and p=0.168 with the alternative approach (PA). Therefore, the model is considered acceptable for prediction with PO (p>0.05) and unacceptable with the PA (p>0.05).

The maximum anticipated errors or critical errors are $e^*_{SC}=0.365$ kg and $e^{**}_{SC}=0.550$ kg; therefore, the modeler or user of the model specifies a precision "e" ($P(|D|\leq e\geq 0.95)$) such that $e\geq e^*_{SC}=0.365$ kg or $e\geq e^{**}_{SC}=0.550$ kg, then the model will be considered reliable enough in the prediction of the system based on the PO and PA, respectively. Thus, if e=0.5 kg model is considered acceptable for prediction with the PO (e≥0.365 kg) and unacceptable with the PA (e<0.550 kg), which coincides with the statistical decisions obtained with the hypothesis testing approach.

The CI of 1-α'=95% for ε_{SC} is 0.353 kg <ε_{SC}< 0.575 kg, so there is a confidence of 1-α' = 95% that the point of the |W| distribution which is under 1-α = 95% of the absolute errors ($|d-\bar{d}|$) is located somewhere in the interval (0.353 kg-0.575 kg), i.e., with a probability of 95% the magnitude of the forecast error will not exceed 0.575 kg.

With a confidence of 95% of the critical error with the PA and the CI for ε_{SC}, a minimum prediction error of 0.550 kg should be allowed, and it would not exceed 0.575 kg. Hence, the validated model could be used to predict the daily AWG of cattle, although it will require an adjustment based on the presence of SC to increase the accuracy of its forecasts.

DISCUSSION

The extensions presented for validating models in the presence of SC are applied without the model being modified in its structure, since the type of bias on the available data (z_i, y_i) is eliminated through the values of the bias (d_i), i.e. $d_i-\bar{d}$ is the correction for SC.

Identifying the type of bias would allow improving the structure and evaluation of the model through testing its structure and the data and methods used in all the processes for its construction and validation. For Barrales et al (3), detecting the presence of bias in the model would allow the user to identify in his model the causes that produce

corrige el SC mediante $w_i=d_i-\bar{d}$. Con $\alpha=\alpha'=0.05$ y e=0.5 kg como el máximo error que admite el modelador o usuario del modelo, se tiene: a) $P(|D|\leq 0.5kg)\geq 0.95$, b) $\sigma_D^2 \leq (e^2/\chi^2_{1,1-\alpha})=0.065$, c) $V^{SC}=25.220$ y nivel de significación p=0.832 con el planteamiento original (PO) y p=0.168 con el alternativo (PA). Por lo tanto, el modelo es considerado aceptable para predicción con el PO (p>0.05) y no lo es con el PA (p>0.05).

Los errores máximos anticipados o errores críticos son $e^*_{SC}=0.365$ kg y $e^{**}_{SC}=0.550$ kg , por lo tanto, si el modelador o usuario del modelo especifica una precisión "e" ($P(|D|\leq e)\geq 0.95$) tal que $e\geq e^*_{SC}=0.365$ kg o $e\geq e^{**}_{SC}=0.550$ kg, entonces el modelo será considerado suficientemente confiable en predicción del sistema con base en el PO y PA respectivamente. Así, si e=0.5 kg el modelo es considerado aceptable para predicción con el PO (e≥0.365 kg) y no lo es con el PA (e<0.550 kg), la cual coincide con las decisiones estadísticas obtenidas con el enfoque de pruebas de hipótesis.

El IC del 1-α'=95% para ε_{SC} es 0.353 kg <ε_{SC}<0.575 kg, por lo que se tiene confianza en un 1-α'=95% de que el punto de la distribución |W| que tiene debajo el 1-α=95% de los errores absolutos ($|d-\bar{d}|$) está localizado en alguna parte en el intervalo (0.353 kg–0.575 kg), es decir, con probabilidad del 95% la magnitud del error de predicción no sobrepasa 0.575 kg.

Con 95% de confianza, del error crítico con el PA y el IC para ε_{SC}, tendría que permitirse como mínimo 0.550 kg de error de predicción y éste no sobrepasaría 0.575 kg. Así, el modelo validado podría usarse para predecir la GPP diaria de bovinos, aunque requerirá un ajuste con base a la presencia de SC para incrementar la exactitud en sus pronósticos.

DISCUSIÓN

Las extensiones presentadas para validar modelos en presencia de SC se aplican sin que el modelo sea modificado en su estructura, ya que se elimina el tipo de sesgo a los datos disponibles (z_i, y_i) a través de los valores del sesgo (d_i), es decir, $d_i-\bar{d}$ es la corrección por SC.

Identificar el tipo de sesgo permitiría mejorar la estructura y evaluación del modelo a través de cuestionar desde su estructura hasta los datos y métodos empleados en todos los procesos de su construcción y validación. Para Barrales et al (3) detectar en el modelo la presencia de sesgo permitiría al usuario identificar en su modelo la o las causas que lo producen y corregir deficiencias

it and correct any deficiencies in the forecasting behavior, what would lead to decrease discrepancies between what has been estimated by the model and the values provided by the real scenario, allowing the conclusions obtained about the reliability of the model to meet the objectives established. On the other hand, Tedeschi (5) indicates that the identification and acceptance of inaccuracies in a model are step towards the evolution of a more accurate and reliable model.

The statistical test and critical error with the alternative approach require a value higher than the maximum permissible error (e) than in the original approach to infer that the model is acceptable. This can be observed, for example, with critical errors, because if $\alpha'=0.5$ then $e^*=e^{**}$ and if $\alpha'<0.5$ then $e^*<e^{**}$. Thus, for a value $\alpha'<0.5$ and "e" such that $e^*<e<e^{**}$, it follows that the model is adequate using the critical error of the original approach ($e>e^*$) and not the alternative ($e<e^{**}$). Reynolds (9) indicates that the statistical test for the alternative approach is more conservative and probably preferable to the original approach by more users of the model who need to be reasonably sure that the model will meet their accuracy requirements. The drawback of applying the original approach is the ambiguity that arises by not rejecting the null hypothesis (H_0), since what can be inferred is that the data do not provide sufficient evidence to reject it and that the statement established in H_0 will not be accepted. In addition, the research hypothesis is proposed in the alternative hypothesis.

Validation by means of the critical errors or the confidence limits approach comes down to calculate the maximum anticipated error or critical error, where the modeler or user decides if the model is acceptable in the prediction of the system, by comparing the critical error with the accuracy required (e) under the values α and α' specified in advance. This involves a sound understanding of the system by the modeler or the user of the model to establish the maximum permissible error (e). Barrales et al (3) indicate that conceptually, the maximum allowable error and the critical error represent the same, but with the difference that the first is established a priori by the modeler whereas the second is calculated ex-post.

The estimated CI of 100(1-α')% for the quantile 1-α of the distribution of errors once the SC is corrected, means a confidence of 100(1-α')% in that the point of the distribution of errors under 100(1-α)% of the absolute errors is located somewhere in such interval and allows determining, with a certain probability, an upper limit for the magnitude of the prediction error. The CI approach to determine if the model is acceptable to predict, according to

en el comportamiento predictivo, lo que llevaría a disminuir las discrepancias entre lo estimado por el modelo y los valores proporcionados por el escenario real, permitiendo que las conclusiones que se obtengan acerca de la confiabilidad del modelo cumplan con los objetivos establecidos. Por su parte Tedeschi (5), señala que la identificación y aceptación de inexactitudes de un modelo es un paso hacia la evolución de un modelo más exacto y de mayor confianza.

La prueba estadística y error crítico con el planteamiento alternativo requieren un valor más grande del máximo error admisible (e) que el planteamiento original para inferir que el modelo es aceptable. Esto puede observarse, por ejemplo, con los errores críticos, ya que si $\alpha'=0.5$ entonces $e^*=e^{**}$ y si $\alpha'<0.5$ se tiene que $e^*<e^{**}$. Así, para un valor $\alpha'<0.5$ y "e" tal que $e^*<e<e^{**}$, se sigue que el modelo es adecuado empleando el error crítico del planteamiento original ($e>e^*$) y no lo es con el alternativo ($e<e^{**}$). Reynolds (9) señala que la prueba estadística para el planteamiento alternativo es más conservadora y probablemente preferible al planteamiento original por más usuarios del modelo quienes necesitan estar razonablemente seguros que el modelo cumplirá con sus requerimientos de exactitud. El inconveniente de aplicar el planteamiento original es la ambigüedad que se presenta al no rechazar a la hipótesis nula (H_0), ya que lo que se puede inferir es que los datos no proporcionan suficiente evidencia para rechazarla y no que se acepte la declaración establecida en H_0. Además, de que la hipótesis de investigación se plantea en la hipótesis alternativa.

La validación por medio de los errores críticos o de enfoque de límites de confianza se reduce a calcular el error máximo anticipado o error crítico, en donde el modelador o usuario decide si el modelo es aceptable en predicción del sistema, al comparar el error crítico con la exactitud requerida (e) bajo los valores α y α' especificados con anticipación. Lo anterior implica una buena comprensión del sistema por parte del modelador o usuario del modelo para establecer el máximo error admisible (e). Barrales et al (3) señalan que conceptualmente el máximo error admisible y el error crítico representan lo mismo, pero con la diferencia de que el primero se establece *a priori* por el modelador, mientras que el segundo se calcula *a posteriori*.

El IC estimado del 100(1-α')% para el cuantil 1-α de la distribución de los errores una vez que se corrige el SC, significa que se tiene confianza en un 100(1-α')% que el punto de la distribución de los errores que tiene debajo el 100(1-α)% de los errores absolutos está localizado en alguna parte de dicho intervalo y permite determinar, con cierta probabilidad, una cota superior para la magnitud del error de predicción. El enfoque de IC para determinar si el modelo es aceptable

requirements of the modeler or the user of the model, is more informative than the hypotheses tests for the same purpose, since they provide the range of possible values for the parameter under consideration and are not as categorical as hypothesis testing.

Moreover, Barrales et al (3) indicate that the confidence limits approach is widely applied, but requires the modeler to be aware of the variability associated with the particular application, in order to allow him to decide, with respect to the magnitude of the limit error (maximum permissible error) in accordance with the accuracy needed for the purposes of the model.

It should be noted that in the validation of a model, it is recommended to use various methods, for example, measures of deviation together with graphical methods (6) and simple linear regression techniques to assess whether a model is accurate and precise (5).

In conclusion, the validation of models based on Freese's approach and those presented in this article, constitute a statistical method consisting of hypothesis testing and confidence intervals to determine if the outputs of the model are sufficiently close to the observed values in the actual system. The method allows analyzing data from models without bias, constant or proportional bias in their forecasts without modifying the structure of the model.

The confidence interval estimate for the quantile $1-\alpha$ of the distribution of errors, once the constant bias has been corrected, allows determining an upper bound for the magnitude of the prediction error and use it to evaluate the evolution of the model in the prediction of the system.

The confidence intervals approach to determine if the model is acceptable to predict is more informative than the hypothesis tests for the same purpose; in the sense that it allows to know the maximum allowable error of prediction ex post. Therefore, the confidence intervals approach is recommended.

Acknowledgements

To the Undersecretary's Office of Higher Education and Scientific Research, Programme for the Improvement of Teacher Training of the Secretary's Office of Public Education of Mexico for financing this research. To UCMEXUS for the support in the postdoctoral work of Ph.D. Luis Vargas-Villamil in the Department of Animal Science of the University of California, Davis.

para predecir, según requerimientos del modelador o usuario del modelo, es más informativo que las pruebas de hipótesis para el mismo propósito, ya que proporcionan la amplitud de los posibles valores del parámetro en consideración y no son tan categóricos como las pruebas de hipótesis.

Por otra parte Barrales et al (3) indican que el enfoque de límites de confianza, es de mayor aplicabilidad, pero requiere por parte del modelador un conocimiento de las variabilidades asociadas a la particular aplicación, de manera que le permita decidir, con respecto a la magnitud del error límite (máximo error admisible) de acuerdo con la exactitud necesitada para los fines del modelo.

Cabe señalar que en la validación de un modelo es recomendable utilizar varios métodos, por ejemplo, medidas de desviación conjuntamente con métodos gráficos (6) y la técnica de regresión lineal simple para evaluar si un modelo es exacto y preciso (5).

En conclusión la validación de modelos, basada en el planteamiento de Freese y los presentados en este artículo, constituyen un método estadístico formado por pruebas de hipótesis e intervalos de confianza para determinar si las salidas del modelo están suficientemente próximas a los valores observados del sistema real. El método permite analizar datos provenientes de modelos sin sesgo, sesgo constante o proporcional en sus pronósticos sin modificar la estructura del modelo.

El intervalo de confianza estimado para el cuantil $1-\alpha$ de la distribución de los errores una vez que se corrige el sesgo constante, permite determinar una cota superior para la magnitud del error de predicción y usarla para evaluar la evolución del modelo en predicción del sistema.

El enfoque de intervalos de confianza para determinar si el modelo es aceptable para predecir es más informativo que las pruebas de hipótesis para el mismo propósito; en el sentido de que permite *a posteriori* conocer el máximo error admisible de predicción. Por lo que se recomienda el enfoque de intervalos de confianza.

Agradecimientos

A la Subsecretaría de Educación Superior e Investigación Científica, Programa de Mejoramiento del Profesorado de la Secretaría de Educación Pública de México por el financiamiento a esta investigación. A UCMEXUS por el apoyo en el trabajo post-doctoral del Dr. Luis Vargas-Villamil en el Departamento de Ciencia Animal de la Universidad de California, Davis.

REFERENCES

1. Oberkampf WL, Trucano TG. Verification and validation in computational fluid dynamics. Prog Aerosp Sci 2002; 38(3):209-72.

2. Halachmi I, Edan Y, Moallem U, Maltz E. Predicting feed intake of the individual dairy cow. J Dairy Sci 2004; 87(7):2254-67.

3. Barrales VL, Peña R, Fernández R. Model validation: an applied approach. Agric Tech (Chile) 2004; 64:66-73.

4. Mayer DG, Butler DG. Statistical Validation. Ecol Model 1993; 68(1-2):21-32.

5. Tedeschi LO. Assessment of the adequacy of mathematical models. Agr Syst 2006; 89(2-3):225-47.

6. Medina-Peralta S, Vargas-Villamil L, Navarro-Alberto J, Canul-Pech C, Peraza-Romero S. Comparación de medidas de desviación para validar modelos sin sesgo, sesgo constante o proporcional. Univ Cienc 2010; 26(3):255-63.

7. Freese F. Testing accuracy. For Sci 1960; 6:139-45.

8. Rennie JC, Wiant HVJ. Modification of Freese's chi-square test of accuracy. Note BLM Denver Colorado: USDI Bureau of Land Management; 1978.

9. Reynolds MR. Estimating the Error in Model Predictions. For Sci 1984; 30(2):454-69.

10. Medina PS. Validación de modelos mecanísticos basada en la prueba ji-cuadrada de Freese, su modificación y extensión. Montecillo, Mexico: Colegio de Postgraduados; 2006.

11. Casella G., Berger RL. Statistical Inference, Pacific Grove CA USA. USA: Duxbury Thompson Learning; 2002.