رقم البحث (20)

DUMMIES VARIABLES IN PRODUCTIVE AND REPRODUCTIVE REGRESSION FUNCTION OF MILK PRODUCTION

BY

Fouda [a], M. M., Tag El Din [b], S. M., Abo El-fadl [a], E. A.

[a]: Department of Animal Husbandry, Faculty of Veterinary Medicine, Mansoura Univeristy, Mansoura, Egypt.

[b]: Department of applied statistics, Institute of Statistical Studies & Research, Cairo University, Cairo, Egypt.

ABSTRACT

This study was carried out through field survey of 5000 animal records of 15 farms in different regions of El-Menofia province, El Dakhalia province and Alexandria province (Dina farm) during the period extended from 2005 to 2011 on random samples of dairy production sectors. These sectors were Small, Private and Governmental farmers. Records included dairy breeds as Balady cattle (Local breed), Holstein-Friesian cattle (Exotic breed), Cross-bred cattle (Balady X Friesian) and buffalo. Housing systems in these farms were free stall or tie stall. Data were collected, arranged, summarized and then analyzed statistically using the computer programs **SPSS/PC$^+$ "Version 16" (SPSS, 2004).** Regression functions of productive and reproductive indices affecting milk production with dummy variables of breeds were used to determine the effect of breed on milk production. Use of dummy variables of breeds in productive function enables us to study the effect of breeds on milk production. It also increased the predictive value of the function from (62%) to (66%). The same effect is recorded when use these dummies in reproductive function as it increased the predictive value of the reproductive function from (21%) to (34%).

INTRODUCTION

Regression makes use of the correlation between variables and the notion of a straight line to develop a prediction equation. Once a relationship has been established between two variables the equation can be developed that allow to predict the score of one of the variables,

given the score of the other ***(Munro, 2006)***. The goal of multiple regression analysis is to determine the values of parameters for a function that causes the function to best fit a set of data observations that the researcher provides. Regression can be used for prediction, inference, hypothesis testing, and modeling of causal relationships (***Lindley, 1987, Fox, 1997; Richard, 2004 and Freedman, 2005).***

Sometime it is necessary to include a categorical variable with more than two levels in a multiple regression prediction model, so, additional steps are needed to insure that the results are interpretable. These steps include recoding the categorical variable into a number of separate, dictomous variables. This recoding is called "dummy variable". We use dummy variables to incorporate qualitative variables into a regression analysis. Dummy variable is one that takes the values 0 or 1 to indicate the absence or presence of some categorical effect that may be expected to shift the outcome. Adding these variables to a linear regression model will always increase the unadjusted R^2 value of the function.

Multiple regression with dummy variables of the breed was used to determine the effect of the breed on milk production of cattle and buffaloes. Multiple regression analysis is an advanced statistical technique that uses more than one predictor, or independent variables, to examine the effects on a single outcome, or dependent variable **(Levine et al., 1997**

The work of this study was done for fitting some model using different regression function with dummy variables which represented qualitative characters.

MATERIAL AND METHODS:

2.1. Data collection and trait studies:

This study was carried out through field survey of 5000 animal records of 15 farms in different regions of El-Menofia province, El Dakhalia province and Alexandria province (Dina farm) during the period extended from 2005 to 2011 on random samples of dairy production sectors. These sectors were Small, Private and Governmental farmers. Small farms represented about 47.3% of the sample, private farms represented 28.3% of the sample whereas; governmental farms represented 24.4% of the sample. Records included dairy breeds as Balady cattle (local breed) which represented 32.12% of the sample, Holstein-Friesian cattle (exotic breed) which represented 24.04% of the sample, Cross-bred cattle (Balady X Friesian) which represented 13.8% of the sample and Buffaloes which represented 30.04% of the sample. Housing systems in these farms were free stall or tie stall.

The data were collected from field survey by two methods:

A. From the farm records which were available in the investigated dairy farms of study area.

B. From the structured questionnaires established by the researcher in accordance with objectives of this study and were admitted to the dairy holders and managers during the time of interview.

The collected data (raw data) were from milk production records and the accompanying reproduction records. The data were classified into:-

Reproductive data which include Breed categories (Balady, Cross-bred, Friesian cattle) and (Buffaloes that was considered as one breed), total milk per year (milk production) / Kg, days in milk or (lactation length/ day), total veterinary supervision cost/ L.E and Feed costs per year /L.E. Reproductive data which include Age at first calving / (day), calving interval /(day) and days open /(day).

2.2. Statistical analysis:

Multiple-regression with dummy variables of breeds was used to construct best fit function of milk production by using its productive and reproductive indices. Data were collected, arranged, summarized and then analyzed statistically using the computer programs **SPSS/PC$^+$ "version 16" (SPSS, 2004).** All the productive and reproductive parameters affecting the dairy production were calculated and statistically analyzed for animals by using regression functions of productive and reproductive indices affecting milk production with dummy variables of breeds to determine the effect of breed on milk production.

Statistical analyses methods:

Firstly homogeneity tests were done to test homogeneity of the data and it included:

1. Hartly´s Test = $\frac{\text{Maximum variance of all the group}}{\text{Summation of all the group variances}}$

Then data were analyzed using relevant statistical methods of data analysis, including regression model.

Regression analysis is a collective name for techniques for the modeling and analysis of numerical data consisting of values of dependent variables (also called response variable or measurement) and of one or more independent variables (also, called explanatory variable).

The parameters are estimated so as to give a "best fit" of the data. Most commonly the best fit is evaluated by using the least square method.

Underlying assumptions:

Classical assumptions for regression analysis include:

1. The sample must be representative of the population.
2. The error is assumed to be a random variable with a mean of zero conditional on the explanatory variables.
3. The independent variables are error-free.
4. The predictors must be linearly independent.

Dummy variable is one that takes the values 0 or 1 to indicate the absence or presence of some categorical effect that may be expected to shift the outcome **(Draper and smith, 1998).**

Calculating number of dummy variable used in equation:

Number of dummy variables used = $r - 1$

(r = number of categories in the studied categorical variables). Group which is not used in calculation is called (omitted group).

Number of dummy variables used in the study = $4\text{-}1 = 3$

(Reference group) here are buffaloes.

2.3. Regression equation with dummy variables being used in the study:

a. Equation of productive indices affecting milk production:

$$\textbf{Log } Y_i = \log b_0 + \log b_1 X_{1i} + \log b_2 X_{2i} + \log b_3 X_3$$

b. Effect of dummy variables of the breed on productive indices affecting milk production:

$$\textbf{Log } Y_i = \log b_0 + \log b_1 X_{1i} + \log b_2 X_{2i} + \log b_3 X_{3i} + \log b_4 d_{1i} + \log b_5 d_{2i} + \log b_6 d_{3i}$$

Where;

X_1= Days in milk X_2 = Feed cost X_3 = Veterinary cost

d_1= 1 if Balady breed d_2 = 1 if Cross bred breed d_3 = 1 if Frisian breed

3. Equation of reproductive indices affecting milk production:

$$\text{Log } Y_i = \log b_0 + \log b_4 X_{4i} + \log b_5 X_{5i} + \log b_6 X_{6i}$$

4. Effect of dummy variables of the breed on reproductive indices affecting milk production:

$$\text{Log } Y_i = \log b_0 + \log b_4 X_{4i} + \log b_5 X_{5i} + \log b_6 X_{6i} + \log b_7 d_{1i} + \log b_8 d_{2i} + \log b_9 d_{31}.$$

X_1=Days open X_2 = Calving interval X_3= Age at first calving

d_1= 1 if Balady breed d_2 = 1 if Cross bred breed d_3 = 1 if Frisian breed

3. Results and Discussion:

Production functions of milk production were estimated in two forms; linear and logarithmic form. But, the logarithmic function was more common and accurate as R^2 was higher in logarithmic form than linear function. They demonstrated the different regression functions and coefficient between milk yield as dependant variable and the different productive and reproductive resources affecting milk yield as independent variables.

Comparison between the outputs of the logarithmic production functions were done to determine the best accurate function, which used to describe the relationships between milk yield and productive resources.

3.1. The logarithmic Productive function of milk production:

It was used to describe the relation between total milk production (Y) as dependant variable and productive efficiency indices (days in milk, feed cost, veterinary cost as independent variable).

Function	**Log Y = -1.194 + 1.35 LogX_1 + 0.169 LogX_2 + 0.341 LogX_3**
t	(11.29)** (28.49)** (22.85)** (30.95)**
F	1368.85**
R^2	0.62
Adjusted R^2	0.61

** Significant at (P<0.01) * Significant at (P<0.05)

X_1 = Days in milk X_2 = Feed cost X_3 = Veterinary cost

The results revealed that the logarithmic production function was highly significant (P<0.01), and about 62 % from the changes in milk yield were attributed to the changes in production efficiency indices. This result was obtained after testing data by **stepwise methods** which indicate that there was no autocorrelation between all variable being used in the regression equation.

3.2. The logarithmic Productive function of milk production with dummies of breed:

It was used to describe the relation between total milk production (Y) as dependant variable and productive efficiency indices (days in milk, feed cost, veterinary cost and dummy variables of breeds (**d1** for Balady breed**, d2** for cross bred and **d3** for Friesian breed) as independent variable.

Function	**Log Y = -0.296+ 1.12 LogX_1 + 0.158 LogX_2 + 0.226 LogX_3+ -0.14 d_1+**
T	**0.010 d_2 - 0.012 d_3**
	(2.59)** (23.14)** (21.77)** (14.29)** (14.24)** (1.075) (1.90)*
F	813.66**
R^2	0.66
Adjusted R^2	0.65

** Significant at (P<0.01) * Significant at (P<0.05)

X_1= Days in milk **$X_{2=}$** = Feed cost **$X_{3=}$** = veterinary cost

d_1= 1 if Balady breed **d_2** = 1 if Cross bred breed **d_3** = 1if Frisian breed

The results revealed that the logarithmic production function was highly significant (P<0.01), and about 66 % from the changes in milk yield were attributed to the changes in production efficiency indices. So, adding dummies of breed led to increase the predictive value of the function from (62%) to (66%). This agrees with **(Back and Winsborough, 1966)** who enumerate that dummy variables can be combined in the regression equation with continuous variables or other dummy variables representing background or previous experimental treatment and lead to increasing predictive value of regression function.

3.3. The logarithmic reproductive function of milk production:

It was used to describe the relation between milk production (Y) as dependant variable and reproductive efficiency indices (age at first calving, days open and calving interval), and the best function as follow:

Function	Log Y = 1.55 + 0.39 LogX$_4$ + 0.733 LogX$_5$ - 0.558 LogX$_6$
t	(6.19)** (11.50)** (6.23)** (15.28)**
F	447.19**
R^2	0.21
Adjusted R^2	0.20

** Significant at (P<0.01)

X_4 = Days open X_5 = Calving interval X_6= Age at first calving

The results revealed that the logarithmic production function was highly significant (P<0.01), and about 21 % from the changes in cow's life time milk yield were attributed to the changes in reproductive efficiency indices. This result was obtained after testing data by **stepwise methods** which indicate that there was no autocorrelation between all variable being used in the equation.

3.4. The logarithmic reproductive function of milk production with dummies of breed

It was used to describe the relation between milk production (Y) as dependant variable and reproductive efficiency indices (age at first calving, days open, calving interval and dummy variables of breeds **d1** for Balady breed**, d2** for crossbred and **d3** for Friesian breed as independent variables and the best function as follow:

Function	Log Y = 1.37 + 0.078 LogX$_4$ + 0.917 LogX$_5$ - 0.312 LogX$_6$ - 0.167 d_1 + 0.0 d_2 - 0.031 d_3
t	(5.99)** (2.41)** (8.51)** (8.50)** (25.70)** (0.019) (4.70)**
F	436.39**
R^2	0.34
Adjusted R^2	0.33

** Significant at (P<0.01) * Significant at (P<0.05)

X_4 = Days open X_5= Calving interval $X_{6=}$ = age at first calving

d_1= 1 if Balady breed d_2 = 1 if Cross bred breed d_3 = 1 if Frisian breed

The results revealed that the logarithmic production function was highly significant (P<0.01), and about 34 % from the changes in cow's life time milk yield were attributed to the changes in reproductive efficiency indices. So, adding dummies of breed led to increase the predictive value of the function from (21%) to (34%).

4. Summary and Conclusion

This study was carried out through field survey of 5000 animal records in different regions of El-Menofia province, El Dakhalia province and Dina farm during the period extended from 2005 to 2011 on random samples of dairy production sectors to evaluate the factors affecting the productive and reproductive efficiency of dairy cattle and Buffaloes under Egyptian conditions. These sectors were Small, Private and Governmental farmers. The dairy breeds included in this study were Balady cow (local breed), Holstein-Friesian (exotic breed), Cross-bred (Balady X Friesian) and Buffaloes. The collected data were summarized, classified then analyzed separately using computer statistical program **(SPSS, 2004)**. The results of this study are summarized in various sub-sections as follow.

1- Use of dummy variables of breeds in productive function enables us to study the effect of the breeds on milk production. It also increased the predictive value of the function from (62%) to (66%).

2- Use of dummy variables of breeds in reproductive function enables us to study the effect of the breeds on milk production of different animal species. It also increased the predictive value of the function from (21%) to (34%).

REFERENCE

Back, K.W., Winsborough, H.H. (1966): The use of multivariate regression using dummy variables in social-physiological experiment, journal of social sciences, 5(3): 77-79.

Draper, N.R. and Smith, H. (1998): Applied Regression Analysis. (3rd Edition), London, UK, Wiley.

Fox, J. (1997): Applied regression analysis, linear models and related methods.sage

Freedman, A.D. (2005): Statistical Models: Theory and Practice. Cambridge University, UK, Press.

Levine, D.M., Berenson, M.L. and Stephan, D. (1997): Statistics for mangers using Microsoft Excel. 3rd edition, Prentice Hall Publication.

Lindely, D.V. (1987): regression and correlation analysis, new palgrave: adictionary of Economics, v.4, pp.120-23.

Munro, B.H. (2006): Statistical methods for health care research. (5th Edition), Pp. (4, 199-204,259-267), ISBN: 0-7817-4840-2.

Richard, A.B. (2004): Regression Analysis: A Constructive Critique. Thousand Oaks, California, USA, Sage Publications.

SPSS, (2004): Statistical Package for Social Sciences. Release 16.0 version. SPSS Inc. USA.

الملخص العربي

استخدام متغيرات وهمية فى الدالة الانتاجية والتناسلية لانتاج الحليب

محمد محمد فوده - سيد مشعال السيد تاج الدين- إيمان أحمد أبو الفضل

قسم الرعاية وتنمية الثروة الحيوانية (أحصاء حيوي)

كلية الطب البيطري – جامعة المنصورة

أجريت هذه الدراسة من خلال المسح الميداني لسجلات ٥٠٠٠ حيوان في مناطق مختلفة من محافظة المنوفية ، محافظة الدقهلية ومحافظة الاسكندرية (مزرعة دينا) خلال الفترة الممتدة من عام (٢٠٠٥-٢٠١١) على عينات عشوائية من قطاعات إنتاج الحليب. وشملت هذه القطاعات مزارع صغار المزارعين، المزارع الخاصة والحكومية. وكانت السلالات المستخدمة في هذه الدراسة: بلدي (سلالة محلية)، الفريزيان هولشتاين (سلالة غربية)، الهجين (بلدي X فريزيان) والجاموس. وذلك لتقييم العوامل التي تؤثر على الكفاءة الإنتاجية والتناسلية في أبقار الحلوب والجاموس. تم جمع البيانات المستخدمة في هذه الدراسة من سجلات دقيقة والتي تتوفر في مزارع الألبان او عن طريق استبيان منظم خلال المسح الميداني. وشملت تلك البيانات العوامل الإنتاجية التي تؤثر على إنتاج الحليب مثل (طول موسم الحليب و تكلفة العليقة والتكلفة الكلية للرعاية البيطرية) . و أيضا على العوامل التناسلية التي تؤثر على إنتاج الحليب مثل (الأيام المفتوحة والعمر عند أول ولادة والفترة بين ولادتين). وقد لخصت البيانات التي تم جمعها وتصنيفها ثم حللت بشكل منفصل باستخدام البرنامج الإحصائي (SPSS، ١٦). تم تحليل البيانات من خلال أساليب إحصائية متعددة بدءا من التحليل الوصفي للبيانات من خلال عرض المتوسطات والأخطاء المعيارية للمتوسطات وصولا لأساليب توثيق البيانات من خلال النماذج الإحصائية المختلفة نموذج الا نحدار اللوغارتيمى لإنتاج الحليب مع إدخال متغيرات وهمية للسلالة لدراسة تأثير السلالة على الدالة الإنتاجية والتناسلية لإنتاج الحليب وقد أوضحت الدراسة إن استخدام تلك العناصر الوهمية للسلالة قد أدى إلى أيضا إلى زيادة القيمة التنبؤية لكل من الدالة الإنتاجية وكذلك الدالة التناسلية لإنتاج الحليب كما أنها قد أدت إلى زيادة دقة الدوال في وصفه للعلاقة بين إنتاج الحليب والعناصر التناسلية والإنتاجية المختلفة التي تؤثر على إنتاج الحليب