

Using the Maple System to Evaluate the Efficiency of a Regression Model

Mirzakarimov Ergashboy Mirzaboyevich

Acting Associate Professor, Fergana Polytechnic Institute, Fergana, Uzbekistan
ergashboy.mir@mail.ru

Ergashev Muhammadrashidkhon Ilhomjon ugli

Assistant, Fergana Polytechnic Institute, Fergana, Uzbekistan
rasheed.off@mail.ru

Annotation:

At present, any serious statistical calculations, as a rule, are carried out on computers and, above all, on personal computers. This investigation reveals the quality and efficiency of mathematical models based on the experimental results of solving engineering and economic problems using the Maple program, as well as their significance in the analysis and decision-making using digital methods [1,2,3,7,8,9,10,11].

ARTICLE INFO

Article history:

Received 18 Mar 2022

Revised form 15 Apr 2022

Accepted 14 May 2022

Key words: regression, criterion, model, adequate, approximation, value, Maple, sec, add, vector.

INTRODUCTION. Mathematical model-linear regression equation of experimental results of regression analysis in statistical data processing using the least-squares method and evaluation of the quality, adequacy, and value of regression coefficients of these models were considered [5,8,9]. In solving these problems, the model is based on the formula for approximation of the average error in the assessment of quality, the adequacy of the model-regression equation from Fisher's F-criterion, and the value of regression coefficients from the student's criterion. In solving all these problems, methods and programs in the Maple system were used [1,2,3,7,8,9,10,11].

MAIN PART. The results of the experiments to determine the relationship between the number of products sold as a result of observations and the amount x spent on its advertising should be as follows.

Table with 2 rows (x, y) and 9 columns of numerical values.

To determine the linear regression equation of this relationship, we construct the following system based on the method of small squares [12]:

System of equations for linear regression coefficients a and b.

From the system of equations, we find the unknowns a and b as follows:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad b = \frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (2)$$

Based on the findings a and b , we write the regression equation as follows [4]:

$$\tilde{y} = ax + b.$$

Statistical evaluation of the found model-linear regression equation.

1. Model quality assessment. An average approximation error is determined to assess model quality. We determine the average deviation (approximation) of the calculated data from the actual data by the following formula:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \cdot 100\%.$$

The smaller these deviations are, the closer the values found are to the theoretical values. This indicates the quality-efficiency of the model. The fact that the average error is in the range of 8-10%, or does not exceed 12-15% rest, determines the quality of the model.

2. Assessment of model adequacy. We determine the adequacy of the regression equation on the basis of Fisher's F -criterion by the following formula:

$$F = \frac{\bar{S}_y^2}{\bar{S}_{gold}^2}$$

where $\bar{S}_y^2 = \frac{1}{k_1} \left(\sum_{i=1}^n y_i - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right)$ is total variance, $\bar{S}_{gold}^2 = \frac{1}{k_2} \sum_{i=1}^n (y_i - \tilde{y})^2$ is residual variance. The

regression equation \bar{y} determines how many times the experimental results are better than the average. We find it from the ratio of the above variances. The value of this criterion F is compared with the value of the tab $F(a, k_1, k_2)$ in the application table, where a is value, $k_1 = n-1$, $k_2 = n-m-1$ degrees of freedom, n is the number of experiments, m is the number of parameters corresponding to x .

In this case:

-the result of the calculation is $F > F_{tab}$, the regression equation found is adequate (statistical value);

-if $F < F_{tab}$, the regression equation is not adequate.

3. Assessment of the significance of regression coefficients.

According to the student criterion, the coefficients a and the values of b are their random errors m_a and We evaluate on the basis of comparison with m_b :

$$t_a = \frac{a}{m_a}, \quad t_b = \frac{b}{m_b}$$

$$m_a = \frac{S_{gold}}{\sigma_x \sqrt{n}}, m_b = \frac{1}{n\sigma_x} S_{gold} \sqrt{\sum_{i=1}^n x_i^2}$$

where is the $\bar{S}_{gold}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \tilde{y})^2$ residual variance and $\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2}$ the standard deviation. These t_a, t_b s are subject to the Student's distribution according to the degree of freedom $v=n-2$. From the distribution table constructed for the student criterion, we find the critical value as $t_{kr} = t(a, v)$:

- 1) If $t_a > t_{kr}$, the parameter a statistically valuable, otherwise not valuable;
- 2) if $t_b > t_{kr}$, parameter b statistically valuable, otherwise not valuable.

Now, in finding the solution to the problem of determining the quality, adequacy, value of the coefficients of the linear regression equation (model) found on the basis of the table compiled on $n = 8$ observations, directly below Maple program.

Maple software

The straight-line connection is based on the above rule:

> restart; with (stats):

> Digits: = 5:

Experimental results:

> x1:=1.5: x2:=4: x3:=5: x4:=7: x5:=8.5: x6:=10:

x7:=11: x8:=12.5:

> y1:=5: y2:=4.5: y3:=7: y4:=6.5: y5:=9.5: y6:=9:

y7:=11: y8:=9:

Random quantities:

> X:=Vector ([x1,x2,x3,x4,x5,x6,x7,x8],

datatype = float):

> Y:=Vector([y1,y2,y3,y4,y5,y6,y7,y8],

datatype = float):

Determination of linear coefficients and equations of random variables:

> n: = 8:

> SX: = add (X [i], i = 1..n);SX := 59.500

> SY: = add (Y [i], i = 1..n);SY := 61.500

> SXX: = add (X [i] ^ 2, i = 1..n);SXX := 541.75

> SYY: = add (Y [i] ^ 2, i = 1..n);SYY := 509.75

> SXY: = add (X [i] * Y [i], i = 1..n);SXY := 510.25

Determining the coefficients of the straight-line connection equation (2) :

> ab:=solve({a*SX+n*b=SY, a*SXX+b*SX=SXY}, {a,b});

$ab := \{a = 0.53260, b = 3.7263\}$

Determine the equation of a straight line connection :

> $yt:=ab[1]*x+ab[2];$

$yt := xa + b = 0.53260x + 3.7263$

> $a:=.532; b:=3.726;a := 0.532 \quad b := 3.726$

Straight line values and sum:

> $Yt:={seq(a*X[i]+b, i = 1..n)};$

$Yt := \{4.5240, 5.854, 6.386, 7.450, 8.2480, 9.046, 9.578, 10.376\}$

> $SXY2:=add((Y[i]-a*X[i]-b)^2, i=1..n) ;$

$SXY2 := 8.8245$

Average deviations and sum of experimental values from a straight line connection:

> $A:=seq(abs((Y[i]-a*X[i]-b)/Y[i]), i=1..n):$

$A:=evalf(%,4)*100;$

$A := 9.49900, 30.1400, 8.73100, 14.6700, 13.1300, 0.570000, 12.8700, 15.3600$

> $SA:=add(abs((Y[i]-a*X[i]-b)/Y[i]), i=1..n)*100;$

$SA := 104.90$

We create a table to determine the values corresponding to the following variables and their sum:

> $print("X[i],Y[i],X[i]^2,Y[i]^2,X[i]*Y[i],Yt=a*X[i]+b, (Y[i]-Yt)^2,A[i]*%");$

$"X[i], Y[i], X[i]^2, Y[i]^2, X[i]*Y[i], Yt=a*X[i]+b, (Y[i]-Yt)^2, A[i]*%"$

> $u:=array(1..9,1..8):$

> for i to n do for j to n do

$u[i,1]:=X[i]; u[i,2]:=Y[i]; u[i,3]:=X[i]^2;$

$u[i,4]:=Y[i]^2; u[i,5]:=X[i]*Y[i]; u[i,6]:=a*X[i]+b;$

$u[i,7]:=(Y[i]-a*X[i]-b)^2; u[i,8]:=A [i];$

$u[n+1,1]:=SX; u[n+1,2]:=SY; u[n,6]:=a*X[i]+b;$

$u[n+1,6]:='-' ; u[n+1,3]:=SXX; u[n+1,4]:=SYY;$

$u[n+1,5]:=SXY; u[n+1,7]:=SXY2; u[n+1,8]:=SA;$

od; od;

> $evalf(evalm(u),4);$

1.5000	5.	2.2500	25.	7.5000	4.5250	0.22558	9.4990
4.	4.5000	16.	20.250	18.	5.8563	1.8396	30.140
5.	7.	25.	49.	35.	6.3888	0.37356	8.7310
7.	6.5000	49.	42.250	45.500	7.4538	0.90973	14.670
8.5000	9.5000	72.250	90.250	80.750	8.2526	1.5561	13.130
10.	9.	100.	81.	90.	9.0513	0.0026317	0.57000
11.	11.	121.	121.	121.	9.5838	2.0056	12.870
12.500	9.	156.25	81.	112.50	10.383	1.9114	15.360
59.500	61.500	541.75	509.75	510.25		8.8242	104.98

We use the values of these variables and the sum of the last row to make a linear relationship in the statistical estimates:

1) The effectiveness of the found link:

> **At:=add(abs((Y[k]-a*X[k]-b)/Y[k]),k=1..n)/n;**

At := 0.13112

> **At:=At*100;At := 13.112**

Mean deviation of values found experimentally with regression correlation: 13.12%.

Error is not greater than 12-15%, the structured link model is a good modulus and allows the calculations to be evaluated (predicted) in advance.

2) Assess the adequacy of the connection:

n-number of experiments and k_1 , k_2 degrees of freedom:

> **n:=8.; m:=1;**

> **k1:=n-1; k2:=n-m-1; k1 := 7. k2 := 6.**

General variance:

> **S2y:=(SYY-SY^2/n)/k1; S2y := 5.2814**

Residual dispersion:

> **S2qol:=add((Y[k]-a*X[k]-b)^2, k=1..n)/k2;**

S2qol := 1.4708

F-Fisher Criterion:

> **F:=S2y/S2qol; F := 3.5908**

According to the Fisher table $F_{tab}(a, k_1, k_2) = F_{tab}(0.05, 7, 6) = 4.21$

> **Ftab:=4.21: #F <Ftab linear connection is adequate**

3) Evaluation of the value of regression coefficients:

Find the random error of the coefficients:

> **sigma[x]:=sqrt(SXX/n-(SX/n)^2); $\sigma_x := 3.5218$**

> **ma:=S2qol/(sigma[x]*sqrt(n)); ma := 0.14766**

```
> mb:=S2qol*sqrt(SXX)/(n*sigma[x]); mb := 1.2151
```

We evaluate the value of the regression coefficients found by their random error according to the Student Criterion:

```
> a: = . 5325984252; b: = 3.726299213;
```

```
a := 0.5325984252 b := 3.726299213
```

```
> ta:=a/ma; ta := 3.6069
```

```
> tb:=b/mb; tb := 3.0667
```

From the distribution table (appendix) according to the Student's criterion, we obtain the critical value: $t_{kr} = t(a, v) = t(a, n - 2) = t(0.05, 6) = 2.45$

In this case, we see that the coefficients a and b are statistically significant, since $t_a > t_{kr}$, $t_b > t_{kr}$:

```
> tkr:=2.45:
```

```
> ta-tkr > 0; 0 < 1.1569
```

```
> tb-tkr > 0; 0 < 0.6167
```

We construct a graph of this linear connection model.

```
> restart;
```

```
> with (stats): with (plots):
```

```
> x1:=1.5: x2:=4: x3:=5: x4:=7: x5:=8.5: x6:=10: x7:=11: x8:=12.5:
```

```
> y1:=5: y2:=4.5: y3:=7: y4:=6.5: y5:=9.5: y6:=9: y7:=11: y8:=9:
```

```
> r2:= rhs(fit[leastsquare[[x,y],y=a*x+b,{a,b}]]([[x1, x2,x3,x4,x5,x6,x7,x8],[y1,y2,y3,y4,y5,y6,y7,y8]]));
```

```
r2 := 0.5325984252 x + 3.726299213
```

```
> Plot ([r2,[[x1,y1], [x2,y2], [x3,y3], [x4,y4], [x5,y5], [x6,y6], [x7,y7],[x8,y8]]], x=1..13, 4..12, style=[line,point], color=[blue,red], thickness=3, symbol=BOX, symbolsize=30); (Figure 1.)
```

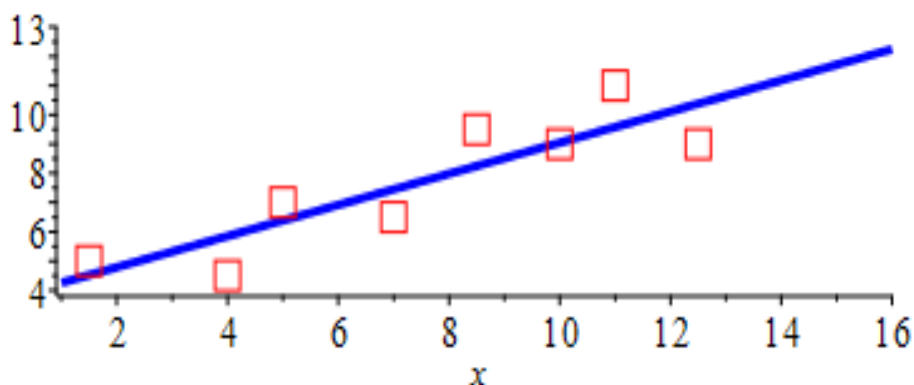


Figure 1.

CONCLUSION. Hence, the linear relation model found by the small squares method is completely adequate to Fisher's F-test, and we find that all its coefficients are valuable. We have seen that accurate, fast, and high-quality results can be obtained using the Maple program to determine the structure and

effectiveness of a model found to draw conclusions and make decisions based on the results of a given experiment.

LIST OF REFERENCES.

1. Abdushukurov A.A., Azlarov T.A., Jamirzaev A.A.,(2003),«Ehtimollar nazariyasi va matematik statistikasidan misol va masalalar to'plam». Toshkent «Universitet».
2. Кремер Н. Ш., (2001), «Теория вероятностей и математическая статистика». Учебное пособие. Москва.
3. Матросов А.В. (2001). Maple 6. Решение задач высшей математики и механики. СПб.: БХВ-Петербург.
4. Mirzakarimov E.M. (2014,2015). Maple dasturi yordamida Oliy matematika masalalarini yechish. 1,2,3q, T.: Adabiyot uchqunlari.
5. Mirzakarimov, E.M., (2021). Maple tizimida to'g'ri to'rtburchakli membraning erkin tebranishini aniqlash. *Scientific Bulletin of Namangan State University* 1(8), 3-9.
6. Mirzakarimov, E.M., (2022). To'g'riburchakli membraning erkin tebranishini boshlang'ich shartlar bo'yicha aniqlashda maple tizimidan foydalanish. *Scientific Bulletin of Namangan State University* 1(1), 53-61.
7. Mirzakarimov, E.M., & Fayzullaev, J.S. (2020). Improving the quality and efficiency of teaching by developing students* mathematical competence using the animation method of adding vectors to the plane using the maple system. *Scientific Bulletin of Namangan State University*, 2(9), 336-342.
8. Mirzakarimov, E.M., & Faizullaev, J.I. (2019). Method of teaching the integration of information and educational technologies in a heterogeneous parabolic equation. *Scientific Bulletin of Namangan State University*, 1(5), 13-17.
9. Mirzaboevich, M.E., & ugli, E. M.I. (2021). Using Maple Programs in Higher Mathematics Triangle Problem Constructed on Vectors in Space. *Central asian journal of mathematical theory and computer sciences*, 2(11), 44-50. Retrieved from <http://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/123>
10. Мирзабоевич, М.Э., & Исмоилджонович, Ф.Д. (2021). Выполнять Линейные Операции Над Векторами В Пространстве В Системе Maple. *Central asian journal of mathematical theory and computer sciences*, 2(12), 10-16. <http://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/137>
11. Мирзабоевич, М.Э., (2022). Использовать Систему Maple Для Определения Свободных Колебаний Прямоугольной Мембраны При Начальных Условиях. *Central Asian Journal of Mathematical Theory And Computer Sciences*, 3(1), 9-18. <https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/148>
12. Mirzakarimov E.M. (2018). Oliy matematika fanidan laboratoriya ishlarini Maple dasturida bajarish. T.: Tafakkur bo'stoni.