MSC 62P30; 62-04 **DOI: 10.14529/mmp200303**

STATISTICAL ANALYSIS MODULE FOR WEIGHT DESIGN OF AIRCRAFT ELEMENTS

***A.I. Kibzun*[1], *A.S. Shalaev*[1], *V.M. Azanov*[1], *A.N. Ignatov*[1]**
[1]Moscow Aviation Institute, Moscow, Russian Federation
E-mails: kibzun@mail.ru, anton.s.shalaev@gmail.com, azanov59@gmail.com, alexei.ignatov1@gmail.com

The concept of a statistical analysis module for weight design of aircraft elements (for predicting weight characteristics of one or another aircraft elements) is proposed. Models, methods to construct single-point estimates of the predicted characteristic, quality criteria of constructed models are considered. Two approaches to the confidence estimation of the predicted characteristic are proposed. First approach is based on the assumption that errors at predicting are caused by inaccurate identification of the deterministic part of the predicted characteristic behavior. The second one is based on the assumption that the deterministic part of the predicted characteristic behavior is identified correctly and errors at predicting are caused by inaccuracy in the measurements. The structure, goals of each component of the software package that implements the statistical analysis module is considered in details. Based on the real data the problem of predicting the take-off mass of an empty equipped airliner depending on maximum pay load and the maximum flight distance at maximum pay load is given. By this problem applicability of considered models and methods is demonstrated.

Keywords: weight design; aircraft; statistical analysis; software.

Introduction

Within the framework of the complex project of Moscow Aviation Institute "Creating training and research system for weight design of the aircraft" a statistical analysis module is created for weight design of the aircraft. At initial stages of aircraft creation, as part of weight design, engineers of the aircraft solve various problems including predicting the maximum take-off weight of the aircraft, the mass of the empty aircraft depending on various factors, for example, the maximum pay load, and maximum flight distance at maximum pay load; prediction of wing mass from the maximum take-off weight, wing area, wing span, and other factors. In each of these problems various factors influence on predicted value. In this regard a certain general mathematical tool is needed: various models, estimation methods that allow to evaluate weight characteristics of one or another element. We propose such a tool in this paper. In the manufacturing process at the aviation industry, the construction and implementation of the statistical analysis module is important problem, since it will allow the formation of high-quality predictions for decision-makers at early stages of aircraft creation (in conditions of multifactorial uncertainty) and thereby improve quality of the design decision.

Among publications devoted to the estimation of various dependences in weight design of aircraft we distinguish among others the following. The work [1] presents an extensive review of various statistical and empirical formulas for predicting various weight characteristics of the aircraft. In addition to [1] we also highlight [2–6]. The paper [2] considers the problem of estimating fuel consumption which is solved using decision trees. In [3] various estimates of the aircraft mass based on least squares methods are proposed, and an algorithm for their convolution into a single random variable characterizing initial mass of the aircraft is given. The paper [4] provides a detailed review of publications on the mass of the aircraft wing. In particular some factor dependencies obtained with the help of least squares method are given. In the framework of [5] an estimate of the aircraft mass is constructed on the basis of least squares method, while in [6] the aircraft mass is proposed to be estimated on the basis of a stochastic dynamic system and constructing a multi-particle filter.

In this paper we propose various models, methods, and criteria for estimating quality of constructed dependencies for weight design of the aircraft elements. In addition to the single-point estimation, two approaches to the confidence estimation of the predicted characteristic are proposed. A meaningful example is considered.

1. Problem Formulation

As we noted in introduction when designing aircraft there are many problems associated with weight design of various elements of the aircraft at a particular design stage. In this regard we consider the problem of weight design in the most general state. At the end of the paper, we will consider the proposed relations using a specific practical example: predicting the take-off mass of an empty equipped airliner depending on maximum pay load and the maximum flight distance at maximum pay load.

First we note that in the ideal case the available set of factors $F_1, F_2, \ldots, F_m$ is enough to find the value of some predicted characteristic Y, that is, there is a dependence

$$Y = f(F_1, F_2, \ldots, F_m),$$

where $f(\cdot)$ is some known function. However, in reality $f(\cdot)$ is unknown and furthermore often the set of factors we have is not enough to explain behavior of the predicted characteristic Y.

2. Models and Methods

Since $f(\cdot)$ is unknown there is a problem about its estimation. Often in practical problems there are the linear model

$$Y = \theta_0 + \theta_1 F_1 + \theta_2 F_2 + \ldots + \theta_m F_m + \varepsilon, \tag{1}$$

and the multiplicative model

$$Y = e^{\theta_0} F_1^{\theta_1} F_2^{\theta_2} \ldots F_m^{\theta_m} \varepsilon, \tag{2}$$

where $\theta_0, \theta_1, \ldots, \theta_m$ are parameters to be estimated and ε is a random variable. These parameters can be confined using some physical constraints $\theta = (\theta_0, \theta_1, \ldots, \theta_m)^{\mathrm{T}} \in \Theta \subset \subset \mathbb{R}^{m+1}$ where Θ is a set of feasible values of $\theta_0, \theta_1, \ldots, \theta_m$. Taking logarithm to the left and right part of (2) from multiplicative model we obtain the linear one

$$\tilde{Y} = \theta_0 + \theta_1 \tilde{F}_1 + \theta_2 \tilde{F}_2 + \ldots + \theta_m \tilde{F}_m + \tilde{\varepsilon}, \tag{3}$$

where $\tilde{Y} = \ln(Y)$, $\tilde{F}_1 = \ln(F_1)$, ..., $\tilde{F}_m = \ln(F_m)$, $\tilde{\varepsilon} = \ln(\varepsilon)$. There are also applicated models of the form

$$Y = g(F_1, F_2, \dots, F_m, \theta_0, \theta_1, \dots, \theta_M) + \varepsilon, \tag{4}$$

where $g(\cdot)$ is some nonlinear function given from physical considerations and $M + 1$ is a quantity of unknown parameters to be estimated.

A distinctive feature of weight design problems from other applied problems of reconstructing physical dependencies is the following. The factors in weight design problems are deterministic that does not allow in particular the use of the often used minimax estimation and sigma-point estimation in which randomness of factors is assumed. Therefore we use traditional methods for recovering factor dependencies: least squares method which is still actively used not only in weight design but also in other applied and fundamental problems [7–9], weighted least squares, quantile regression.

Suppose that we have n observations (aircrafts): y_i is the value of the predicted characteristic Y and f_{ij} is the value of the factor F_j for the i-th observation, $i = \overline{1,n}$. Due to the possible uncertainty of some observations in addition to the classical least squares method it makes sense to consider weighted least squares where each observation has a certain weight $w_i \geq 0$, $i = \overline{1,n}$ and in order to estimate of $\theta_0, \theta_1, \dots, \theta_M$ it is necessary to solve the problem

$$\sum_{i=1}^{n} w_i (y_i - g(f_{i1}, f_{i2}, \dots, f_{im}, \theta_0, \theta_1, \dots, \theta_M))^2 \to \min_{\theta \in \Theta}.$$

Note that in the framework of the developed software package an user is invited to assign each observation to one of five observation groups: reliable ($w_i = 1$), likely reliable ($w_i = 0{,}75$), neutral ($w_i = 0{,}5$), subject to doubts ($w_i = 0{,}25$), unreliable ($w_i = 0$). Such gradation is caused by the fact that data on the weight characteristics of aircraft elements are often a trade secret therefore aircraft manufacturing companies may distort data on the weight characteristics of various elements. By default all observations are assumed to be neutral and therefore the classic least squares coincides with the weighted ones.

One more method to estimate the unknown parameters $\theta_0, \theta_1, \dots, \theta_M$ is quantile regression [10, 11]. For searching of estimates it is necessary to solve the problem

$$\sum_{y_i \geq g(f_{i1}, f_{i2}, \dots, f_{im}, \theta_0, \theta_1, \dots, \theta_M)} \alpha |y_i - g(f_{i1}, f_{i2}, \dots, f_{im}, \theta_0, \theta_1, \dots, \theta_M)| +$$
$$+ \sum_{y_i < g(f_{i1}, f_{i2}, \dots, f_{im}, \theta_0, \theta_1, \dots, \theta_M)} (1 - \alpha) |y_i - g(f_{i1}, f_{i2}, \dots, f_{im}, \theta_0, \theta_1, \dots, \theta_M)| \to \min_{\theta \in \Theta}. \tag{5}$$

Taking $\alpha = 1/2$ in (5), we obtain least absolute deviations method which is less sensitive to emissions than least squares method. Note that if the function $g(\cdot)$ is not bounded by $\theta \in \Theta$ then taking $\alpha = 0$ or $\alpha = 1$ we can construct the regression line enveloping all observations. Namely when $\alpha = 0$ we get that the predicted value of weight characteristic of the aircraft element at each fixed set of the factors f_i is not more than the the exact value y_i; on the other hand, at $\alpha = 1$ the predicted value of weight characteristic of the aircraft element is not less than the exact value, $i = \overline{1,n}$.

3. Criteria for Model Choice

Let $\theta_0^*, \theta_1^*, \ldots, \theta_M^*$ be estimates of $\theta_0, \theta_1, \ldots, \theta_M$ obtained using one of the methods mentioned above in this article.

To check the quality of the constructed model we can use the adjusted criterion R^2 widely used in practice [12–14]

$$R^2 = 1 - \frac{n-1}{n-M-1} \frac{\sum\limits_{i=1}^{n} (y_i - g(f_{i1}, f_{i2}, \ldots, f_{im}, \theta_0^*, \theta_1^*, \ldots, \theta_M^*))^2}{\sum\limits_{i=1}^{n} (y_i - \overline{y}_n)^2},$$

where $\overline{y}_n = \sum\limits_{i=1}^{n} y_i/n$. In practical problems it is assumed that more large coefficient R^2 and more close to 1 leads to proposed model describes a certain physical/social/economic process better. However a single criterion of R^2 is not enough to draw conclusion on quality of the constructed model. So for example in the case when the predicted characteristic has a wide range of values and at the same time there is some physical dependence of the predicted characteristic on several factors consideration of any one factor allows to obtain an acceptable value of the criterion R^2 but at the same time the predicted values may be significantly far from the exact values.

The deviation of the predicted values from the exact values can be calculated using the mean absolute error

$$\Delta = \frac{1}{n} \sum_{i=1}^{n} |y_i - \tilde{y}_i|,$$

where $\tilde{y}_i$ is a predicted value obtained by one or another method. However, the use of Δ as a quality criterion is also fraught with difficulties. Proximity of the criterion Δ to zero in predicting one characteristic and a deviation from zero in another does not necessarily mean that the prediction is constructed more accurately in the first problem. This effect is caused by the fact that if the values of the predicted characteristic are close to some small number then the value of the criterion Δ a priori is close to zero. If the values of the predicted characteristic are in the billions then the value of Δ is huge.

The latter drawback is eliminated by the use of the average relative error

$$\delta = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \tilde{y}_i|}{y_i}.$$

4. Confidence Estimation

The criteria mentioned in the previous section make it possible to assess the applicability built models and its quality. However, on new data (i.e. observations that are neither in a training sample, nor in a testing sample), when there are only values of factors and the exact value of the predicted characteristic is unknown, these criteria do not allow to assess how close the predicted value is to the exact one. For this purpose, we construct confidence intervals for the predicted characteristic Y using the estimates $\theta_0^*, \theta_1^*, \ldots, \theta_M^*$ of $\theta_0, \theta_1, \ldots, \theta_M$ obtained by classical least square method. We describe two possible ways of constructing confidence intervals.

The first way is to find the estimates $\theta_0^*, \theta_1^*, \ldots, \theta_M^*$ of unknown parameters $\theta_0, \theta_1, \ldots, \theta_M$ in one or another model and subsequent estimation of the distribution law of the errors ε. Due to the fact that the number of observations on the aircraft is small, test of hypothesis on the form of the distribution law of a random variable ε is complicated, we suppose that the distribution is normal and log-normal in (4) and (2), respectively.

Consider (4), where the error ε is normally distributed and $\theta_0^N, \theta_1^N, \ldots, \theta_M^N$ is an optimal solution obtained by the least squares method. Then using maximum likelihood estimation we obtain $\varepsilon \sim \mathcal{N}(\hat{m}, \hat{D})$, where

$$\hat{\varepsilon}_i = y_i - g(f_{i1}, f_{i2}, \ldots, f_{im}, \theta_0^N, \theta_1^N, \ldots, \theta_M^N), \;\; \hat{m} = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i \;\; \hat{D} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\varepsilon}_i - \hat{m})^2.$$

Since

$$\mathcal{P}\left(c + \hat{m} - u_{1-\beta/2}\sqrt{\hat{D}} \le c + \varepsilon \le c + \hat{m} + u_{1-\beta/2}\sqrt{\hat{D}}\right) = 1 - \beta,$$

where c is a some constant and $u_{1-\beta/2}$ is $1-\beta/2$ – quantile of standard normal distribution, then $1-\beta$ confidence interval for the predicted characteristic Y is

$$[g(F_1, F_2, \ldots, F_m, \theta_0^N, \theta_1^N, \ldots, \theta_M^N) + \hat{m} - u_{1-\beta/2}\sqrt{\hat{D}},$$
$$g(F_1, F_2, \ldots, F_m, \theta_0^N, \theta_1^N, \ldots, \theta_M^N) + \hat{m} + u_{1-\beta/2}\sqrt{\hat{D}}].$$

Consider model (2), where the error ε is log-normally distributed, we make transformation (3), $\theta_0^L, \theta_1^L, \ldots, \theta_M^L$ are given by the least squares solution. Again using maximum likelihood estimation we obtain $\tilde{\varepsilon} \sim \mathcal{N}(\tilde{m}, \tilde{D})$, where

$$\tilde{\varepsilon}_i = \ln(y_i) - \theta_0^L - \sum_{k=1}^{m} \theta_k^L \ln(f_{ik}), \;\; \tilde{m} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i \;\; \tilde{D} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{\varepsilon}_i - \tilde{m})^2.$$

Therefore, the random variable ε is log-normally distributed with the parameters $\tilde{m}$ and $\tilde{D}$. Since

$$\mathcal{P}(cz_{\beta/2} \le c\varepsilon \le cz_{1-\beta/2}) = 1 - \beta,$$

where $z_{\beta/2}$ is $\beta/2$ – quantile and $z_{1-\beta/2}$ is $1-\beta/2$ – quantile of log-normal distribution with the parameters $\tilde{m}$ and $\tilde{D}$, $1-\beta$ confidence interval for the predicted characteristic Y is

$$[z_{\beta/2} e^{\theta_0^L} F_1^{\theta_1^L} F_2^{\theta_2^L} \ldots F_m^{\theta_m^L}, z_{1-\beta/2} e^{\theta_0^L} F_1^{\theta_1^L} F_2^{\theta_2^L} \ldots F_m^{\theta_m^L}].$$

The second approach consists in using statistical properties obtained by the least squares method and assumption that the deterministic part in model (1) is identified correctly, i.e. the predicted characteristic Y depends on only these factors and exactly linearly on parameters. The error ε is supposed to be normal with zero expectation. In this case, having error is explained by inaccuracy in the measurement of factor values and predicted characteristic.

Consider model (1), where $\theta_0^R, \theta_1^R, \ldots, \theta_m^R$ are given by the least squares method solution. For model (1) according to [15] we obtain the following $1-\beta$ confidence interval for useful signal, i.e. for deterministic part in model (1)

$$[\theta_0^R + \theta_1^R F_1 + \ldots + \theta_m^R F_m - u_{1-\beta/2}\sqrt{\sigma^2 f^{\mathrm{T}} W^{-1} f},$$
$$\theta_0^R + \theta_1^R F_1 + \ldots + \theta_m^R F_m + u_{1-\beta/2}\sqrt{\sigma^2 f^{\mathrm{T}} W^{-1} f}],$$

where

$$W = H^{\mathrm{T}}H, \;\; H = \begin{pmatrix} 1 & f_{11} & f_{12} & \dots & f_{1m} \\ 1 & f_{21} & f_{22} & \dots & f_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & f_{n1} & f_{n2} & \dots & f_{nm} \end{pmatrix}, \;\; f = \begin{pmatrix} 1 \\ F_1 \\ \dots \\ F_m \end{pmatrix},$$

and σ^2 is the variance of the random variable ε. Since value of the variance σ^2 is unknown, we use its estimate

$$\sigma^2 \approx \frac{1}{n-m-1}\sum_{i=1}^{n}(y_i - \theta_0^R - \sum_{k=1}^{m}\theta_k^R f_{ik})^2.$$

It is also possible to construct the confidence interval for multiplicative model (2) using statistical properties of least squares method. To this end, suppose that in (3) $\tilde{\varepsilon}$ is normally distributed with the parameters 0 and $\sigma^2_{\tilde{\varepsilon}}$. According to [15] note

$$\tilde{L} = \theta_0 + \theta_1\tilde{F}_1 + \ldots + \theta_m\tilde{F}_m - (\theta_0^L + \theta_1^L\tilde{F}_1 + \ldots + \theta_m^L\tilde{F}_m) \sim \mathcal{N}(0, \sigma^2_{\tilde{\varepsilon}}\tilde{f}^{\mathrm{T}}\tilde{W}^{-1}\tilde{f}), \quad (6)$$

where

$$\tilde{W} = \tilde{H}^{\mathrm{T}}\tilde{H}, \;\; \tilde{H} = \begin{pmatrix} 1 & \ln(f_{11}) & \ln(f_{12}) & \dots & \ln(f_{1m}) \\ 1 & \ln(f_{21}) & \ln(f_{22}) & \dots & \ln(f_{2m}) \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \ln(f_{n1}) & \ln(f_{n2}) & \dots & \ln(f_{nm}) \end{pmatrix}, \;\; \tilde{f} = \begin{pmatrix} 1 \\ \ln(F_1) \\ \dots \\ \ln(F_m) \end{pmatrix}.$$

Taking exponent from left and right part of equality (6) and noting that the exponent of normal distribution is log-normal distribution we obtain that the confidence interval for useful signal, i.e. for deterministic part in model (2) is

$$[z_{\beta/2}e^{\theta_0^L}F_1^{\theta_1^L}F_2^{\theta_2^L}\ldots F_m^{\theta_m^L}, z_{1-\beta/2}e^{\theta_0^L}F_1^{\theta_1^L}F_2^{\theta_2^L}\ldots F_m^{\theta_m^L}],$$

where $z_{\beta/2}$ and $z_{1-\beta/2}$ are $\beta/2$ and $1-\beta/2$ – quantiles of log-normal distribution with the parameters 0 and $\sigma^2_{\tilde{\varepsilon}}\tilde{f}^{\mathrm{T}}\tilde{W}^{-1}\tilde{f}$. Since $\sigma^2_{\tilde{\varepsilon}}$ is unknown, then we use its estimate

$$\sigma^2_{\tilde{\varepsilon}} \approx \frac{1}{n-m-1}\sum_{i=1}^{n}(\ln(y_i) - \theta_0^L - \sum_{k=1}^{m}\theta_k^L\ln(f_{ik}))^2.$$

The fundamental difference between the first approach to constructing confidence intervals and the second is that in the second approach the error ε is used to explain measurement inaccuracies, while in the first approach it is used to explain inaccuracies in the model. In this case, mathematically, the second approach is not completely correct because of the possible incorrect identification of the model and replacement of the variance of the error with its estimate. Note that the confidence interval for the linear model in the first approach has the same length in contrast with the confidence interval for the linear model in the second approach in which the length of the interval depends on a specific set of factor values. Note that the variable length of the confidence interval is more logical because for large values (for example, a million) of the predicted characteristic the length of the confidence interval should obviously be different than for values close to zero.

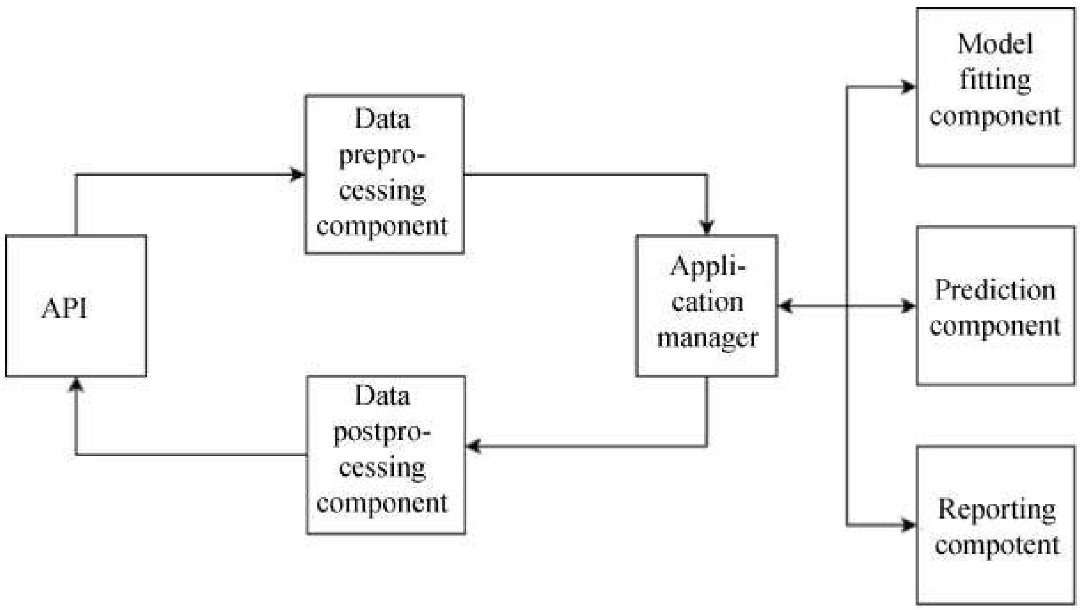
5. Structure of Software

The software package is a web server application with a user and software interface and database. The application software shell is a container (Docker technology) with the system and software dependencies installed in it. The latter fact allows to easily transfer the application from one server to another without installing the necessary system and software dependencies preinstalled in the container in advance and manage the versioning of libraries, environments, structural and system dependencies.

The user interface of the application allows

- create a new task by submitting the training sample (csv, xls, xlsx formats) and selecting factors, target variable, and objects of observation;
- calculate weight characteristics in an existing task using the “best” (by specific criterion) model;
- create a new model by submitting the training sample to the input (csv, xls, xlsx formats) and selecting factors, a target variable, objects of observation and put the model in a collection;
- calculate weight characteristics in an existing task using an arbitrarily selected model from the collection;
- save calculation results in a database;
- save and publish a calculation report.

The program interface (API) allows the application to interact with third-party resources by transmitting the relevant data through specified requests. A backend application is designed using the modular architecture shown in figure.



Block diagram of the backend application

The modular architecture allows testing user functionality of the application. We describe the purpose of the individual components of the backend application.

1. The data preprocessing component is intended for

   1.1) translation of specified API requests into a set of commands to the Application manager;

   1.2) functional data preprocessing for their subsequent transfer to the Model fitting, Prediction, Reporting components in accordance with a task trigger.

2. Application manager is intended for formation of the task trigger, i.e. determining the sequence of tasks by Model fitting, Prediction, Reporting components in accordance with the task trigger.

3. Model fitting component is intended for

   3.1) fitting of statistical models using training dataset;

   3.2) transfer and save the fitted model to Prediction, Reporting components.

4. Prediction component is intended for

   4.1) formation of predictions on a target sample;

   4.2) transfer predictions to Reporting component.

5. Reporting component is intended for formation of specified reports of numerical experiments fitting statistical model and making predictions.

6. The data preprocessing component is intended for

   6.1) Formation of specified answers to API requests;

   6.2) Post-processing of functional data obtained during the execution of Model fitting, Prediction, Reporting components in accordance with the task trigger.

The backend application is developed in Python, component testing is implemented using Pytest technology. The application database is developed using open PostgreSQL technology and is a standard relational model.

6. Example

Consider the problem of predicting the take-off mass of an empty equipped airliner (OEW) depending on maximum pay load (MaxPL) and the maximum flight distance at maximum pay load (MaxD). We will us linear model (1) in which every estimated parameter is not less than zero since aircraft mass can not be non-positive, multiplicative model (2) and non-linear Evdokimov's model [1]

$$g(F_1, F_2, \ldots, F_m, \theta_0, \theta_1, \ldots, \theta_M) = \theta_0 \cdot \mathrm{MaxPL} \cdot \mathrm{MaxD} \cdot \left(\frac{1}{\theta_1 (10^{-3} \cdot \mathrm{MaxD} + \theta_2)} + \theta_3 \right).$$

First we present the training sample collected from open sources on the basis of which we will estimate parameters of models.

Table 1

Training sample

Plane model	OEW, kg	MaxPL, kg	MaxD, km	Plane model	OEW, kg	MaxPL, kg	MaxD, km
Il-114	15000	6500	1000	B 737-300	31480	15404	4006
Dash 8 Q100	10433	4082	1889	B 737-600	36378	15558	4076
Dash 8 Q200	10501	4195	1713	B 737-700	37648	17554	3827
Dash 8 Q300	11793	6124	2034	B 737-800	41413	21319	3145
Dash 8 Q400	17186	8670	2522	B 737-900	42901	20738	3472
ATR72	12950	7850	2220	B 737-900ER	44676	23045	4417
ERJ135	11501	4499	2224	B 757-200	62100	21350	5642
ERJ140	11808	5292	2074	B 757-300	64580	30690	4230
ERJ145	12038	5862	1823	A318	38818	15682	3713
CRJ200-200	13730	5411	2491	A319	39725	18674	4023
CRJ200-200LR	13835	6124	3148	A321	48510	26944	4380
CRJ700-700	19731	8528	2655	Tu-204-300	54000	18000	5920
CRJ700-700ER	19731	8528	3209	Tu-204-200	59000	25200	4350
CRJ1000 EL	23179	12156	2761	B 767-200	80127	33271	4263
CRJ1000	23179	12156	2761	B 767-200ER	82377	35557	9082
CRJ1000 ER	23179	12156	3131	B 767-300	86069	40230	4410
ERJ170	21040	9100	3255	B 767-300ER	90011	43799	7395
ERJ175	21620	10200	3088	B 767-400ER	103147	46583	6850
ERJ195	28850	13530	2924	B 777-200LR	145150	63957	14083
ARJ21-700	25000	8935	2200	B 777-300ER	167829	69853	10655
ARJ21-700ER	25000	8935	3700	Il-96M	132400	58000	7600
ARJ21-900	26270	11246	2200	A330-200	117041	53260	7286
An-148-100	25380	9000	1150	A330-300	120132	54868	4929
An-148-200	25380	12000	900	A340-200	125242	47758	10700
CS110	33200	13971	4074	A340-500	170370	61630	12244
CS130	35500	16556	5463	B 747-400	179752	67457	10589
B 717-200	31071	14515	2297	B 747-400Combi	184113	72167	9728
B 737-200	28622	13472	3649	B 747-400ER	184567	67177	11581
B 737-500	31312	15182	3476	A380	270281	90718	11856

Using classical least squares method we obtain results presented in Table 2. As follows from the Table 2 the best accuracy is given by the multiplicative model. However, its accuracy is inadequate. To construct a more accurate prediction it is necessary to take into account other factors for example the diameter of the fuselage, the maximum number of passengers on board. Let us construct the OEW prediction on the test sample comparing this prediction with exact values and also construct 0,95 confidence intervals.

Table 2

Quality of the constructed models

Estimate	R^2	Δ	δ
$\mathrm{OEW} = 2,474 \cdot \mathrm{MaxPL}$	0,967	7162	14,1
$\mathrm{OEW} = 1,414 \cdot \mathrm{MaxPL}^{0,952} \cdot \mathrm{MaxD}^{0,114}$	0,979	5595	9,9
$\mathrm{OEW} = 0,007 \cdot \mathrm{MaxPL} \cdot \mathrm{MaxD} \cdot \left(\frac{1}{64,82(10^{-3} \cdot \mathrm{MaxD} - 2,44)} + 0,035\right)$	0,854	17082	50,8

The confidence interval constructed by the linear model in the first approach regardless of the values of the factors has a length approximately 34,000 kg which is very large therefore this interval is not shown in the Table 3. In the column "Confidence intervals" the confidence intervals are presented in the following way: in 1 approach in the multiplicative model, in 2 approach in the linear model, in 2 approach in the multiplicative model.

Table 3

Verification of the proposed models with a test sample

Plane model	Exact value of OEW	MaxPL	MaxD	Linear model		Multiplicative model		Confidence intervals
				Prediction	Abs./rel. error	Prediction	Abs./rel. error	
B 777-300	157800	66730	6672	165090	7290/4.6	151067	6733/4.3	[118600, 192534] [155787, 174393] [142154, 160584]
ATR42	11250	5450	2100	13483	2233/19.8	12196	946/8.4	[9575, 15544] [9853, 17113] [11477, 12965]
ERJ145XR	12591	5909	2635	14619	2028/16.1	13517	926/7.3	[10612, 17228] [10923, 18314] [12720,14369]
CRJ900-900	21432	10319	2956	25529	4097/19.1	23286	1854/8.7	[18281, 29678] [22256, 28802] [21912,24753]
ERJ190	27720	12720	3563	31469	3749/13.5	29029	1309/4.7	[22790, 36997] [28235, 34703] [27316,30858]
ARJ21-900ER	26770	11246	3300	27823	1053/3.9	25592	1178/4.4	[20092, 32617] [24534, 31112] [24082, 27205]
B 737-400	33189	19427	3340	48062	14873/44.8	43124	9935/29	[33856, 54961] [44981, 51143] [40579,45840]
A320	41345	19756	3605	48876	7531/18.2	44202	2857/6.9	[34703, 56336] [45968, 51784] [41594,46987]
B 777-200	138100	57980	10492	143443	5343/3.9	139145	1045/0.8	[109240, 177339] [137787, 149099] [130935,147910]
A340-600	176364	74636	9847	184649	8285/4.7	175685	679/0.4	[137928, 223910] [177314, 191986] [165320,186753]

As follows from the Table 3 the multiplicative model also gives better results than the linear model on the test sample. Moreover in terms of the relative error large values of the predicted characteristic are better predicted due to the fact that least squares method minimizes the sum of squared residuals and obviously for "large" observations there may also be large residuals. Therefore least squares method tends to draw a regression line closer to "large" observations. This effect wiil be eliminated by consideration of a larger number of factors. A smaller length of confidence intervals is provided by the multiplicative model. Moreover although exact values are not always within such confidence interval the upper limit always covers the exact value. At the same time the upper limit which is rather conservative due to the high level of reliability is often not far from the exact value. There is one abnormal observation of the Boeing 737-400 for which the available set of two factors is obviously not enough to predict.

Conclusion

In this work we proposed the set of models, methods, and quality control criteria for the constructed models for the statistical analysis module for weight design of aircraft elements. Two methods were proposed for constructing confidence intervals for the predicted characteristic (weight of one or another element of the aircraft). The meaningful example based on real data was considered which showed the applicability of factor models and the proposed models and methods for weight design.

References

1. Sheinin V.M., Kozlovsky V.I. *Vesovoe proektirovanie i effektivnost' passazhirskih samoletov. T. 1. Vesovoi raschet samoleta i vesovoe planirovanie* [Weight Design and Airliner Efficiency. Vol. 1. Weight Airliner Evaluation and Weight Planning]. Moscow, Mashinostroenie, 1977. (in Russian)
2. Chenyu Huang, Yixi Xu, Johnson M.E. Statistical Modeling of the Fuel Flow Rate of GA Piston Engine Aircraft Using Flight Operational Data. *Transportation Research Part D: Transport and Environment*, 2017, vol. 53, pp. 50–62. DOI: 10.1016/j.trd.2017.03.023
3. Junzi Sun, Ellerbroek J., Hoekstra J.M. Aircraft Initial Mass Estimation Using Bayesian Inference Method. *Transportation Research Part C: Emerging Technologies*, 2018, vol. 90, pp. 59–73. DOI: 10.1016/j.trc.2018.02.022
4. Dababneha O., Kipouros T. A Review of Aircraft Wing Mass Estimation Methods. *Aerospace Science and Technology*, 2018, vol. 72, pp. 256–266. DOI: 10.1016/j.ast.2017.11.006
5. Alligier R., Gianazza D., Durand N. Learning the Aircraft Mass and Thrust to Improve the Ground-Based Trajectory Prediction of Climbing Flights. *Transportation Research Emerging Technologies*, 2013, vol. 36, pp. 45–60. DOI: 10.1016/j.trc.2013.08.006
6. Junzi Sun, Blom H., Ellerbroek J., Hoekstra J. Aircraft Mass and Thrust Estimation Using Recursive Bayesian Method. *8th International Conference on Research in Air Transportation*, Barcelona, Spain, 2018, 8 p.
7. Shapeev V.P., Bryndin L.S., Belyaev V.A. Solving Elliptic Equations in Polygonal Domains by the Least Squares Collocation Method. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2019, vol. 12, no. 3, pp. 140–152. (in Russian) DOI: 10.14529/mmp190312
8. Livshits M.Y., Nenashev A.V., Pleshivtseva Yu.E. Computational Algorithm for Optimal Control of an Object with Distributed Parameters in a Nonsmooth Area of Final States. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2019, vol. 12, no. 4, pp. 41–51. (in Russian) DOI: 10.14529/mmp190403
9. Sagadeeva M.A. Reconstruction of Observation from Distorted Data for the Optimal Dynamic Measurement Problem. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2019, vol. 12, no. 2, pp. 82–96. (in Russian) DOI: 10.14529/mmp190207

10. Goryainova E.R., Pankov A.R., Platonov E.N. *Prikladnye metody analiza statisticheskih dannyh* [Applied Methods of Statistical Data Analysis]. Moscow, Izd. dom Vysshei shkoly ekonomiki, 2012. (in Russian)

11. Xiang Liu, Saat M.R., Xiao Qin, Christopher P.L. Analysis of U.S. Freight-Train Derailment Severity Using Zero-Truncated Negative Binomial Regression and Quantile Regression. *Accident Analysis and Prevention*, 2013, vol. 59, pp. 87–93. DOI: 10.1016/j.aap.2013.04.039

12. Leri M.M., Pavlov Yu.L. On the Robustness of Configuration Graphs in a Random Environment. *Informatics and Applications*, 2018, vol. 12, no. 2, pp. 2–10. (in Russian) DOI: 10.14357/19922264180201

13. Longfei Chen, Zhirong Liang, Haoye Liu, Shirun Ding, Yanfei Li. Sensitivity Analysis of Fuel Types and Operational Parameters on the Particulate Matter Emissions from an Aviation Piston Engine Burning Heavy Fuels. *Fuel*, 2017, vol. 202, pp. 520–528. DOI: 10.1016/j.fuel.2017.04.052

14. Fama E.F., French K.R. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 1993, vol. 33, pp. 3–56. DOI: 10.1016/0304-405X(93)90023-5

15. Pankov A.R., Platonov E.N. *Praktikum po matematicheskoj statistike* [Mathematical Statistics Tutorial]. Moscow, Izd. MAI, 2006. (in Russian)

Received April 10, 2020

УДК 519.248 **DOI: 10.14529/mmp200303**

МОДУЛЬ СТАТИСТИЧЕСКОГО АНАЛИЗА ДЛЯ ВЕСОВОГО ПРОЕКТИРОВАНИЯ ЭЛЕМЕНТОВ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

***А.И. Кибзун*[1]*, А.С. Шалаев*[1]*, В.М. Азанов*[1]*, А.Н. Игнатов*[1]**
[1]Московский авиационный институт, г. Москва, Российская Федерация

Предлагается концепция модуля статистического анализа для весового проектирования элементов летательных аппаратов (для прогнозирования веса того или иного элемента летательных аппаратов). Рассматриваются модели, методы для построения точечных оценок прогнозируемой характеристики, критерии качества построенных моделей. Предлагаются два подхода к доверительному оцениванию прогнозируемой характеристики. В одном подходе предполагается, что ошибки в прогнозировании вызваны неточной идентификацией детерминированной части поведения прогнозируемой характеристики. В другом подходе предполагается, что детерминированная часть поведения прогнозируемой характеристики идентифицирована верно, а ошибки в прогнозировании вызваны неточностью измерений. Подробно рассматривается структура, задачи каждой из компонент программного комплекса, реализующего модуль статистического анализа. На основе реальных данных рассматривается задача прогнозирования массы пустого снаряженного пассажирского самолета от двух факторов: максимальной коммерческой нагрузки и максимальной дальности при максимальной коммерческой нагрузке, в которой демонстрируется применимость предлагаемых моделей и методов.

Ключевые слова: весовое проектирование; летательные аппараты; статистичеекий анализ; программный комплекс.

Литература

1. Шейнин, В.М. Весовое проектирование и эффективность пассажирских самолетов. Т. 1. Весовой расчет самолета и весовое планирование / В.М. Шейнин, В.И. Козловский. – М.: Машиностроение, 1977.
2. Chenyu Huang. Statistical Modeling of the Fuel Flow Rate of GA Piston Engine Aircraft Using Flight Operational Data / Chenyu Huang, Yixi Xu, M.E. Johnson // Transportation Research Part D: Transport and Environment. – 2017. – V. 53. – P. 50–62.
3. Junzi Sun. Aircraft Initial Mass Estimation Using Bayesian Inference Method / Junzi Sun, J. Ellerbroek, J.M. Hoekstra // Transportation Research Part C: Emerging Technologies. – 2018. – V. 90. – P. 59–73.
4. Dababneha, O. A Review of Aircraft Wing Mass Estimation Methods / O. Dababneha, T. Kipouros // Aerospace Science and Technology. – 2018. – V. 72. – P. 256–266.
5. Alligier, R. Learning the Aircraft Mass and Thrust to Improve the Ground-Based Trajectory Prediction of Climbing Flights / R. Alligier, D. Gianazza, N. Durand // Transportation Research Part C. Emerging Technologies. – 2013. – V. 36. – P. 45–60.
6. Junzi Sun. Aircraft Mass and Thrust Estimation Using Recursive Bayesian Method / Junzi Sun, H. Blom, J. Ellerbroek, J. Hoekstra // 8th International Conference on Research in Air Transportation. – Barcelona, 2018. – 8 p.
7. Шапеев, В.П. Решение эллиптических уравнений в полигональных областях методом коллокации и наименьших квадратов / В.П. Шапеев, Л.С. Брындин, В.А. Беляев // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2019. – Т. 12, № 3. – С. 140–152.
8. Лившиц, М.Ю. Вычислительный алгоритм оптимального управления объектом с распределенными параметрами в негладкой области конечных состояний / М.Ю. Лившиц, А.В. Ненашев, Ю.Э. Плешивцева // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2019. – Т. 12, № 4. – С. 41–51.
9. Сагадеева, М.А. Построение наблюдения для задачи оптимального динамического измерения по искаженным данным / М.А. Сагадеева // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2019. – Т. 12, № 2. – С. 82–96.
10. Горяинова, Е.Р. Прикладные методы анализа статистических данных / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. – М.: Издат. дом Высшей школы экономики, 2012.
11. Xiang Liu. Analysis of U.S. Freight-Train Derailment Severity Using Zero-Truncated Negative Binomial Regression And Quantile Regression / Xiang Liu, M.R. Saat, Xiao Qin, P.L. Christopher // Accident Analysis and Prevention. – 2013. – V. 59. – P. 87–93.
12. Лери, М.М. Об устойчивости конфигурационных графов в случайной среде / М.М. Лери, Ю.Л. Павлов // Информатика и ее применения. – 2018. – Т. 12, № 2. – С. 2–10.
13. Longfei Chen. Sensitivity Analysis of Fuel Types and Operational Parameters on the Particulate Matter Emissions from an Aviation Piston Engine Burning Heavy Fuels / Longfei Chen, Zhirong Liang, Haoye Liu,Shirun Ding, Yanfei Li // Fuel. – 2017. – V. 202. – P. 520–528.
14. Fama, E.F. Common Risk Factors in the Returns on Stocks and Bonds / E.F. Fama. K.R. French // Journal of Financial Economics. – 1993. – V. 33. – P. 3–56.
15. Панков, А.Р. Практикум по математической статистике / А.Р. Панков, Е.Н. Платонов. – М.: Изд-во МАИ, 2006.

Андрей Иванович Кибзун, доктор физико-математических наук, профессор, кафедра «Теория вероятностей и компьютерное моделирование», Московский авиационный институт (г. Москва, Российская Федерация), kibzun@mail.ru.

Антон Сергеевич Шалаев, ведущий специалист научно-исследовательского отдел кафедры «Проектирование и сертификация авиационной техники», Московский авиационный институт (г. Москва, Российская Федерация), anton.s.shalaev@gmail.com.

Валентин Михайлович Азанов, кандидат физико-математических наук, кафедра «Теория вероятностей и компьютерное моделирование», Московский авиационный институт (г. Москва, Российская Федерация), azanov59@gmail.com.

Алексей Николаевич Игнатов, кандидат физико-математических наук, кафедра «Теория вероятностей и компьютерное моделирование», Московский авиационный институт (г. Москва, Российская Федерация), alexei.ignatov1@gmail.com.

Поступила в редакцию 10 апреля 2020 г.