ABSTRACT
The significant part of biomedical research is engaged in the testing of behavior and organism reaction to different impacts due to animal models. The task of effects on motor and cognitive functions related to neurodegenerative diseases is also solved at the Department of Pathophysiology, Faculty of Medicine in Pilsen, Charles University. Possibilities include the effect of embryonic cerebellum survives transplantation, the impact of forced motor activity or the effects of the chemical compound. The correct interpretation of the results and errors elimination of the experiment is therefore very important. The paper presents commonly used statistical methods for motor and cognitive test evaluation, new possibilities for experimental data evaluation (with minimal and rapid intervention in terms of methods), their comparison and the assessment of the errors associated with the evaluation of experimental data from a motor and cognitive testing.

KEY WORDS
Neurodegenerative diseases, motor and cognitive tests, experimental data evaluation, error analysis, biomedical computing

1 Introduction
The purpose of biomedical research is to understand the principles of the human body, to examine the impact of various diseases and to find appropriate methods for preventing diseases. Animals are important in all parts of research, although the use of live animals is just one of three major research methods in medicine and biology. The most frequently used animal models in medical research are still mice. In recent decades, many mice models were developed for laboratory research of neurodegenerative diseases. See [1] for review. A few of these models are used at the Department of Pathophysiology at Faculty of Medicine in Pilsen too, see [2] e.g.

The correct data evaluation and interpretation of results is very important for motor and cognitive function testing, as well as followed motor and spatial learning modeling associated with neurodegenerative diseases. During the performed experiment evaluation, it is necessary to consider laboratory experiment methodology, since a lot of the computer models are validated based on the utilization of laboratory experiments results. The inaccurate evaluation method, experiment methodology or ineligible results interpretation can commit a systematic error. Moreover, uncertainty prevails during the execution of the experiment. It is desirable to target the error analysis because the elimination of mentioned errors is always required.

Due to the demonstration of evaluation options, their comparison and a part of error analysis it is necessary to describe the chosen example of performed experiment. Basics for detailed analysis could be find in [3].

1.1 Experiment description
The chosen performed experiment focused on nerve functions of a mouse strain C57Bl/6 after cerebellar transplantation, cognitive and motor abilities of individuals in concrete. The nerve functions are tested with regard to the influence by training which proceed from the 11th to the 55th day after surgery either swimming in the Morris water maze (MWM) or training on the cylinder.

Functions were tested in several ways: spatial learning in MWM, open-field, motor coordination test on the horizontal bar, on the ladder (head up) and the cylinder (head in the direction of rotation) and epileptic reactions in the followed sequence. Four motor coordination tests were performed before surgery (a priori information) and 4th, 8th and 10th week after the surgery. In one session the trial was repeated four times. The trial was considered as successful when the animal did not fall down within 60 s (the time limit) or if it left the equipment actively (it jumped down). After finishing the coordination tests the spatial learning (as cognitive function test) was repeated for 10 consecutive days using the MWM method [4]. If the mouse did not find the platform within 60 s, it was placed there. On the last day was performed open-field testing, the animal was placed in the center of the square arena and its behavior (distance) was recorded over a 5 minutes period. After finishing the spatial learning process, the epilepsy reaction was recorded.

The first group of the trained and untrained specimens were transplanted piece of tissue, the second group was the...
control sample, the successfull of transplantation was also assessed by embryonic cerebellar graft survive – graft survived or graft not found. Another factor is the gender of the individual and the type of epileptic reaction. The resulting data set includes 88 subjects with transplant of which 44 were untrained, 18 trained by swimming in MWM and 26 trained on a cylinder and 56 without the transplant of which 26 were untrained, 10 trained by swimming in MWM and 20 trained on the cylinder.

1.2 Standard evaluation of experiment

The performed experiments are valuated solely by statistic methods in recent time. The overview of used statistical methods in biology can be found in [5]. The standard evaluation procedure of behavior testing can be found e.g. in [2]. Given that the aim of the performed experiment is the effect of embryonic cerebellar tissue transplantation to nerve functions of trained and untrained mice, at first it is necessary to determine the differences and relationships between a group of untrained and trained mice. More detailed group division will not be addressed in this article because it is focused on models of error and the way of experiments evaluation.

The standard evaluation procedure of motor co-ordination tests is to use mean and standard error of the mean (μ ± SEM) to determine the time of the equipment fall down, the average percentage of test success in each day and visualization of these tests. A chosen demonstration example of motor tests evaluation is testing on the cylinder because this equipment was performed in one training group. A comparison is made for groups defined on the basis of a hypothesis of training effect on motor skills, see Figure 1. The evaluation requires an average time of equipment fall down period, value of active leaving limit set to 60 s.

![Figure 1. Mean and SEM of motor tests success [%] according to the training type.](image)

Mann-Whitney U-test is used for comparing results. It is one of the most widely used nonparametric test of statistical significance to test the hypothesis that two independent random samples come from the same distribution (test of median equality). It should be mentioned that success is tested at the level of significance α = 0.02 and the time of equipment fall at the level of significance α = 0.05.

Using the standard ways of evaluation, it was discovered that statistically significant differences in success are in the second and third test. It is understandable that the difference should not be considered during the first test because the testing was performed before the surgery and training of individuals was in progress afterwards. The most noticeable difference is between the group without training and trained groups.

The probability match in the second day for the group without training and the group trained by swimming in the MWM is \( p = 1.12 \cdot 10^{-4} \), for the group without training and the group trained on the cylinder \( p = 1.31 \cdot 10^{-4} \); in the third day for the group without training and the group trained by swimming in the MWM \( p = 1.46 \cdot 10^{-5} \), the group without training and the group trained on the cylinder \( p = 3.43 \cdot 10^{-6} \). For the time of equipment fall down, the results are similar, but less statistically significant differences can be found in the first and then in as well as the last test. Success values and the time spent on equipment is significantly higher for both trained groups compared to the untrained group. In accordance to the above, it is important to mentioned that in case of data number > 20 is statistics calculation approximated by normal distribution and statistical significance is determined by z-table values. (see [6])

The standard evaluation procedure does not entirely show the interdependence between each trial of the test because of data averaging, success probability during trials, individuals abilities distribution in each test of experiment. This information is important for defining errors or the impact strength of priori information.

The cognitive functions test in the MWM evaluation is by standard uses \( \mu \pm \text{SEM} \) for the latencies of the platform reaching, swimming distance and swimming velocity for each day and visualization of the spatial learning process. Movements of the mice were recorded by the automatic tracking system EthoVision (Noldus company). The standard evaluation of spatial learning is shown on the Figure 2 and 3.

The analysis of variance (ANOVA) is used for the comparison of samples defined on the basis of assumptions regarding influence of training on cognitive abilities. This is typically used along with the Scheffe’ critical value, obtained from the F distribution.

Escape latency and swimming distance are significantly less for the group trained on the cylinder compared with no training group (at 95% significance level since 4th to 6th day of the test, at 90% level of significance the difference is already apparent on the 3rd day) and compared with the group trained by swimming (at 95% significance level already between the 4th to 6th day of the test), demonstrated on Figure 2. Swimming velocity is statistically significantly lower for trained individuals in the swimming group compared with no training in the 3rd and 4th day of test.

Demonstrative comparison is given in Tab. 1, containing the values of escape latency in the 4th test day. The probability of conformity between the individual distribu-
2 Analysis of experiment factors

2.1 Statistical evaluation

For the first view of the experiment, it is appropriate to use another methods of biological statistics. This part focuses on the evaluation view for unbalanced data, i.e. the use of median and interquartile range, addresses the correlation for experiment outputs and the first sight to the influence of individual factors by using ANOVA.

The first significant aspect of the experiment is provided by correlation analysis. The correlation for all 3 motor tests is minimal. It is therefore necessary to evaluate each test separately. In the case of the cognitive test, the correlation between latency, velocity and distance was made solely to verify $s = v \cdot t$. Latency as the output is shown the most suitable for evaluation and fully adequate for learning process description.

When examining the impact of individual factors using ANOVA (testing the hypothesis of means equality between different groups of random select, intergroup and intragroup variability), it was discovered that for all 3 motor tests the most significant factor is training, followed by the transplantation and its success. For the cognitive test, the most significant factor is the training, but the second most significant factor, unlike the motor test, is epileptic reaction. This test had to be made at the beginning of the evaluation since the obtained information can be used during the next step of evaluation.

To present a different approach for evaluation, using the parameters of normal distribution, there are given statistical parameters of motor test on the cylinder presented for the group without training for the 4th test (10th week after surgery) in the Table 2.

<table>
<thead>
<tr>
<th>1 group</th>
<th>2 group</th>
<th>( -2 \cdot \sigma )</th>
<th>odizd ( \mu_1 - \mu_2 )</th>
<th>( 2 \cdot \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>no training</td>
<td>swimming</td>
<td>10.25</td>
<td>1.73</td>
<td>6.79</td>
</tr>
<tr>
<td>no training</td>
<td>cylinder</td>
<td>0.15</td>
<td>9.24</td>
<td>18.13</td>
</tr>
<tr>
<td>swimming</td>
<td>cylinder</td>
<td>0.85</td>
<td>10.97</td>
<td>21.10</td>
</tr>
</tbody>
</table>

The results suggest a large data unevenness, that is why therefore the evaluation rather shows quantile information. It is necessary noted that the selected example is not extremely. The directly measured data of motor test have little informative value therefore it would benefit from percentile sequence where the median is considered the most frequently used percentile standard.

2.2 Use of cumulative distribution function for experimental data evaluation

The use of cumulative distribution function assessed the information about probability of equipment fall down under
that time limit (motor tests), the probability of reaching the island under the time limit or time spent in the center of the maze (MWM) and the probability distribution of achieving a certain distance during open-field testing. The use of the cumulative distribution function for the experiment evaluation is demonstrated for motor tests on the group tested on the cylinder and for cognitive test in MWM with using the data of latency.

The distribution function $F(x)$ of arbitrary random variables $X$ is defined for $x \in \mathbb{R}$ with relation $F(x) = P(X < x)$ [6]. It is a very suitable method for the reconstruction of data and identification of parameters for subsequent modeling.

The usage of experimental distribution function is very convenient when comparing samples and changes in a particular data set. For example, it can be used even if it is required to compare at least two principles of the success evaluation of motor test. The first principle is to evaluate the success during each test and then compute the mean and SEM (see Figure 1), the second is to evaluate the success during each trial of the test. In principle, these two views are incomparable. The advantage of the first approach is partially the elimination of fluctuation during the test, the disadvantage is disregarding the each test dynamics. The second principle respects the dynamics. Because the learning process is evaluated, the dynamics plays an important role – not only between tests but also between each trials (this is noticable during the assessment of the impact of a priori information use e.g., see section 3.2). For a scope of principles comparison, just specify the changes in distribution functions during all 4 test, see Figure 4.

The graph describing the motor test on the cylinder (Figure 5) evaluates the execution of all 4 trials of performed 4 tests because of complying with the dynamics, values of $x = -3, 0 >$ means the values of the priori information (the first test) before training.

Providing that the strongest factor in the experiment (from ANOVA) was determined as the training, it is appropriate to focus on this data specifically. The Figures 5 and 6 show the success rates of the performed tests in each groups according to training type.

**Figure 5.** Success of the motor test on the cylinder.

**Figure 6.** Success of the cognitive test in MWM (latency).

The graph describing the motor test on the cylinder (Figure 5) evaluates the execution of all 4 trials of performed 4 tests because of complying with the dynamics, values of $x = -3, 0 >$ means the values of the priori information (the first test) before training.

From a statistical point of view, figures 5 and 6 clearly show as a statistical evaluation that training has influence crosswise, i.e. the best results in MWM are achieved by individuals trained in motor function and the best results in motor tests are achieved by individuals trained in the maze (but for MWM, the role of the time ismore significant, not...
just success). The distribution function of other tests is only informative, open-field test e.g., because this test is performed only once, one trial after the test in a maze in the 10\textsuperscript{th} week.

2.3 Use of approximation of phase–type distribution for experimental data evaluation

To represent the distribution of the reference data set, it is also possible to use other methods, e.g. to use the phase approximation of the distribution which eliminates the upper limitation of experiments (60 s) and describes the distribution of failure in the case of motor tests and success in the case of the MWM test. The phase–type distribution is selected due to its properties, see [7] e.g. These probability density functions are suitable for describing the stochastic processes having discrete states in continuous time. An important property of phase–type distributions is that any continuous distribution defined on the set of non-negative numbers can be approximated by phase distribution with an arbitrary precision. This feature is demonstrated in [8]. Data fitting concretely uses the Coxian phase–type distribution [9] with 14 states, due to the sufficient combination of deviation from the real data and the computational complexity. It should be noted, in this case the Coxian distribution is not considered as a representation of queueing systems (such as in [10]) but as an instrument for evaluation.

Figure 7 shows the density of the real data and the density function of the selected phase–type approximation (the curve) of escape latency for the group trained by swimming.

3 Error analysis for experiment evaluation methodology

Due to nature of study, it is possible to focus on particular errors associated with the evaluation of experimental data and perform an analysis. The experimental data obtained and evaluated are limited by many factors which caused deviations. Therefore, when interpreting the results of the experiments, the aim is to eliminate errors as much as possible. This section presents a few factors and their influences.

3.1 Consequences of different approach of asset classification

Using of any manner of evaluation, it is possible to commit errors related to the improper way of asset classification for data obtained from performed experiment. If it is desirable to evaluate the effect of transplantation from the exemplary experiment for example, it is necessary to consider the results of the ANOVA – a more significant factor of this experiment is training. This part shows differences caused by direct evaluation of group with transplantation and control group and these two groups in different groups according to training. This is demonstrated on the cognitive test in the MWM for an evaluation procedure using $\mu \pm \sigma$.

Figure 8. Mean and standard deviation of the escape latency for the group with transplantation and control group.

The curves in the Figure 8 do not make it clear, which group is better in cognitive test in terms of how strong the effect of transplantation is. However, taking into account the strongest influence of training, it was discovered that with the group trained on the cylinder, individuals in the control group achieved better results (Figure 9). Groups trained by swimming and without training contain the worst individuals in control group. If the classification system of obtained data is not well-chosen, this information cannot be received.

3.2 Consequences of not using a priori information

For 92 cases of 145 tested subjects, is available a priori information about motor skills in the data set. Therefore, it is suitable to focus only on this group and use the knowledge to the learning process correction based on unbiased motor skills of each group of individuals. To explain the
consequences of not using a priori information, the group trained on the cylinder is selected to determine its effect on motor functions. This is demonstrated for an evaluation procedure using $\mu \pm \sigma$ for motor testing.

Figure 9. Mean and standard deviation of the escape latency for the group trained on the cylinder.

Figure 10. Mean and standard deviation of the time spent on equipment in group trained on the cylinder.

Figure 11. Mean and standard deviation of the time spent on equipment increasing for the group tested on the cylinder.

Priori information can be used several ways such as calculation of the mean of all 4 trials during priori (first) test and all the curves shifts according to their difference on the same level. Another option is to take the value of each trial and as an expression of learning, considered the change of the parameters, see Figure 11.

Considering where priori information is not use, it is impossible to make this correction and information about learning process could be inaccurate or distorted. The motor test on the horizontal bar, for mentioned group, is the most problematic (the sum of increasing for all 12 trials is 61.87), the test on the ladder is the less problematic (the sum of increase is 258.99) and for the test on the cylinder, the sum of increasing is 204.40.

3.3 Errors caused by using different approximation for real data

As mentioned above, the normal distributions are the most common type of distribution found in nature but they are not the only. But this manner of data identification and evaluation is not the best way. This chapter lists selected distributions for the approximation of data, namely the probability density function is given by the formula (more in [6] and [9]):

- normal distribution
  \[
  f(x|\mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
  \]

- exponential distribution
  \[
  f(x|\lambda) = \left\{
  \begin{array}{ll}
  \lambda e^{-\lambda x} & \text{if } x \geq 0 \\
  0 & \text{otherwise}
  \end{array}
  \right.
  \]

- Weibull distribution
  \[
  f(x|a, b) = \left\{
  \begin{array}{ll}
  ba^{-b}x^{b-1} e^{-(\frac{x}{b})^a} & \text{if } x \geq 0 \\
  0 & \text{otherwise}
  \end{array}
  \right.
  \]

- Coxian phase–type distribution
  \[
  f(\tau|S, S_0, \alpha) = \alpha e^{\tau S} S_0,
  \]

where

\[
\alpha = [1, 0, 0, \ldots, 0] \quad S = \begin{bmatrix}
-\lambda_1 & \lambda_1 (1 - \beta_1) & 0 & \ldots & 0 \\
0 & -\lambda_2 & \lambda_2 (1 - \beta_2) & \ldots & 0 \\
0 & 0 & -\lambda_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & -\lambda_n
\end{bmatrix} \\
S_0 = [\beta_1 \lambda_1, \beta_2 \lambda_2, \ldots, \beta_n \lambda_n]^T
\]

where $\lambda_i$ intensity of remaining in state $s_i$ and $\beta_i$ probability of change over state $s_i$ to an absorbing state and these distributions are compared. For the discrepancy between the selected approximation and the experimental distribution function determination, each test uses $\chi^2$ criterion:

\[
\chi^2 = \sum_{i=1}^{60} \frac{(F_B(i) - F_A(i))^2}{F_A(i)},
\]

where $F_A(i)$ are values of the selected distribution function and $F_B(i)$ are values of the experimental distribution function.

The following figures illustrats the $\chi^2$ deviations of approximations for the individual tests. First, errors are displayed for the latency of the cognitive test (Figure 12).

The size of the area under the curves is used for the error results comparison. The best approximation with the lowest error function ($\sum \chi^2 = 30.21$) is the Coxian phase–type. The following are approximations by the Weibull
function with an error value 42.35, by the normal distribution with 42.55 (these two curves are almost overlapping in Figure 12) and the highest error function with an error value 103.50 has the exponential distribution.

Given that the standard procedure of evaluation is based on the normal distribution, Tab. 3 shows the probabilities of conformity in terms of the experimental distribution with the normal distribution in each part of the cognitive test for latency:

Table 3
Probability of conformity in terms of the experimental distribution of latency with the normal distribution (for the level of significance $\alpha = 0.05$)

<table>
<thead>
<tr>
<th>Day</th>
<th>$\chi^2_{Coxian}$</th>
<th>$\chi^2_{Normal}$</th>
<th>$\chi^2_{Exponential}$</th>
<th>$\chi^2_{Weibull}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st day</td>
<td>0.1031</td>
<td>0.0028</td>
<td>0.0012</td>
<td>0.0963</td>
</tr>
<tr>
<td>2nd day</td>
<td>0.0344</td>
<td>0.0420</td>
<td>0.3577</td>
<td>0.0250</td>
</tr>
<tr>
<td>3rd day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12th day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following Figures 13 – 18 shows the errors for the motor tests. Graphs sections are shown for the process of all errors, because they are not obvious from the original graphs. According to the size of the area under the curves, the best approximation with the lowest error function for all 3 motor tests is the Weibull function. For motor test on the cylinder the value of the error function is 0.27, for the ladder it is 0.20 and for the horizontal bar 0.29. The following are: for the test on the cylinder the approximation by the normal distribution with the error value 2.00, by the exponential distribution (2.39) and the highest error function has the Coxian distribution (43.48); for the test on the ladder the approximation by exponential function with the error value 0.57, by the normal distribution (4.53) and the highest error function has the Coxian phase-type distribution 118.51 and for the test on the horizontal bar, following is the normal distribution approximation with the error value 2.08, then the phase–type approximation (4.70) and the highest error function with the error value 4.95 has the exponential distribution.

4 Conclusion

During the course of evaluation in terms of performed laboratory experiments on animal (mice) models, the proper interpretation of the results and errors elimination is very important. When performing the experiment and interpreting results, it is possible to commit a systematic error and an uncertainty is always present. In the paper there were on the chosen experiment connected with neurodegenerative research presented commonly used statistical methods for motor and cognitive test evaluation, suggested few pos-
sibilities of the experimental data evaluation and mentioned errors connected with not well-chosen methods.

For the purpose of evaluation in the medicine branch, I would like to propose a path with less intervention in the standard procedure of evaluation, based on the normal distribution and related tests. This means that for unbalanced data, the median and the interquartile range will be used more like the mean and standard deviation (or SEM), to the methodology of experiment performing always including measurement before the investigation, due to some effects (in the case of not using a priori information is not possible to make some corrections) and utilize other methods of statistics such as correlation and primal information on the impact of factors to individuals by ANOVA or MANOVA.

For a more complex but more precisely accurate evaluation, I suggest a path with more intervention, meaning use of the information obtained directly from the experimental distribution functions (time and success) or for time data identification and results comparison via using the Coxian phase-type distribution for cognitive tests and the Weibull distribution for motor tests.

Finally, it should be noted that during the evaluation, it is necessarily to always contemplate on how to deal with the dynamics of the process. It depends on what is asked, which hypotheses should be confirmed or rejected.

Acknowledgement

This paper has been supported by the Ministry of Education of the Czech Republic under the project No. 1M0567.

References


