

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 26, číslo 4, prosinec 2015

KRÁTKÉ POJEDNÁNÍ O PROBLÉMU ROZVRHOVÁNÍ S PŘEDEPSANÝMI ČASY PRACÍ A NÁHODNÝMI PRVKY

A NOTE ON FIXED INTERVAL SCHEDULING WITH STOCHASTIC ELEMENTS

Martin Branda^{1,3}, Jan Novotný²,
Asmund Olstad¹, Pavel Popela²

Address: ¹Molde University College, P. O. Box 2110, 6402 Molde, Norway

²Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno

³Department of Probability and Math. Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Prague 8

E-mail: branda@karlin.mff.cuni.cz, ynovot35@stud.fme.vutbr.cz,
asmund.olstad@himolde.no, popela@fme.vutbr.cz

Abstrakt: Úloha „Fixed Interval Scheduling“ (FIS) se zabývá nalezením optimálního rozvrhu prací, jejichž počátek a konec zpracování jsou pevné a předem dané. V tomto článku představíme stochastickou verzi FIS úlohy motivovanou logistickou složkou oprav a údržby strojů. Náš výzkum je motivovaný především rozvrhováním v současném ropném průmyslu v Norsku, konkrétně údržbou a opravami příbřežních ropných plošin. Nejprve formulujeme dvoustupňovou úlohu stochastického programování. Dále využijeme modelovací prostředí GAMS. V závěru diskutujeme výpočetní aspekty úlohy a uvedeme výsledky, které otevírají další směry výzkumu.

Klíčová slova: Fixed interval scheduling, FIS, GAMS, stochastické programování a plánování.

Abstract: The Fixed Interval Scheduling (FIS) problem deals with scheduling of jobs to machines, where all the jobs have given fixed starting and finishing times. In this paper, we introduce a stochastic version of the FIS problem, motivated by the logistic area of technical servicing and maintenance. Our research is particularly inspired by the scheduling challenges of today's Norwegian oil-industry sector, namely within offshore oil-platform maintenance. We formulate a two-stage stochastic programming model. The GAMS environment is used to conduct numerical experiments. We discuss the modelling aspects and present computational results which point towards future research.

Keywords: Fixed interval scheduling, FIS, GAMS, stochastic programming and stochastic scheduling.

1. Introduction

Scheduling is concerned with optimal allocation of scarce resources to activities over time. It is an area of operations research characterized by a virtually unlimited collection of problem types [8]. The problems share one common starting point – there is a set of activities J_1, J_2, \dots, J_n ¹ which have to be scheduled. By scheduling an activity, we mean the decision about when it should be processed, and which resources (e.g. machines, teams or material) will be allocated to the activity.

Fixed interval scheduling is a special case of scheduling, where the temporal data are given and the decision is restricted only to the resource allocation. Each activity has a fixed starting and completion time and preemption is not allowed – each activity is processed in the given (fixed) time-interval.

As in every research area, there exist multiple acronyms for the same concepts. Fixed interval scheduling is sometimes referred-to as Fixed Job Scheduling, and sometimes simply as Interval Scheduling. Fixed interval scheduling has grown into a rich research area in the course of time. The survey paper [4] notes several variants of the problem and mentions that e.g. crew scheduling is a special application area of fixed interval scheduling. Within crew scheduling (with Personnel Scheduling, Manpower Planning and Rostering as acronyms), the survey paper [2] from 2004 categorizes 700 different research papers on the subject, while the survey paper [1] from 2013 lists 291 articles since 2004.

Most scheduling problems are NP-hard [3], and scheduling problems in general are among the most difficult to solve. Since the time-aspect of the fixed interval scheduling problem is missing, it partly alleviates the complexity. This allows to expand the complexity of the resource-allocation part, where each resource can be modelled in fine detail (e.g. considering various qualifications or degrees of experience of crew members, etc.).

The most simple form of the fixed interval scheduling problem consists of a single machine and a set of intervals (activities), which may overlap. The goal is to find the biggest (in terms of cardinality) subset of non-overlapping intervals, since the machine may process at most one interval at a time. A greedy algorithm “order by finishing times” solves this problem. This problem is naturally extended to deal with multiple machines or resources. If the objective is to schedule as many activities as possible given the available resources, we talk about Operational Fixed Interval Scheduling [6].

¹When this set is not known in advance and is only revealed in the course of time, we talk about online scheduling.

Another instance of the simplest fixed interval scheduling problem is to schedule all activities using a minimum number of machines.² The number of activities which pairwise overlap is a lower bound for the number of machines needed. The problem can be solved using a “left-edge algorithm” where the jobs are assigned to the machines in order of nondecreasing starting times, using a machine used before whenever possible. This algorithm has complexity $O(n \log n)$. The problem is extended by considering e.g. various machine types and other resource constraints. When the objective is to determine the minimum amount of resources in order to schedule all activities, we talk about Tactical Fixed Interval Scheduling [7].

The extended versions of the above-mentioned problems are usually NP-hard and require specialised exact or heuristic algorithms.

2. Motivation

Our motivation comes from challenges in offshore oil-platform maintenance related to crew scheduling. The situation as described below is relevant to Norway of 2012–2014.

When maintaining oil platforms, crews of engineers and technicians are transported to the offshore platform to carry out a contracted set of tasks, within a limited time period – typically 14 days. The companies which provide the maintenance services are usually responsible for maintaining multiple platforms. They can process several requests simultaneously and it schedules the maintenance activities in advance. There is little flexibility in scheduling the maintenance jobs, as these are determined by strict technical conditions and constraints (and are mostly given by the platform operators, including the time window for performing the service). Each maintenance activity has several requirements regarding the skills of the servicing personnel and the size of the crew. The skills are formalized by means of certificates, which guarantee the ability of their holders to carry out certain tasks.

The service-providing companies then face the decision problem of allocating their engineers and technicians to the planned jobs. The allocation plan must obey the above constraints:

- Job requirements (necessary skills of the personnel).
- Overall schedule (scheduling parallel tasks requires separate crews).
- Availability of the staff (personal calendars including holidays, medical appointments, etc.).
- Working regulations (rest periods, overtime, etc.).

²This problem is sometimes referred-to as Interval Partitioning.

A good allocation has the property that the resources (technicians) are effectively utilized within the legal limits. The scheduling problem can be solved on two levels – tactical and operational, see [6] and [7]:

- *Tactical*: The goal is to determine how many jobs can be accepted to a long-term schedule – given the current resources and the uncertainties involved.
- *Operational*: Given the set of jobs, decide upon the optimal personnel allocation.

The operational problem is an inherent part of the tactical problem, possibly in an approximative or simplified form. The operational problem alone can be used for daily re-planning. Our focus in this paper is mostly on the tactical level.

The uncertainties are present in the following way: If a job takes longer-than-expected time to complete, the operating crew becomes unavailable for the next scheduled job. In this case, if no other crew is available, the company may decide to outsource personnel from the market (work-force pool). A conservative approach in job-acceptance (maintaining large time buffers) will lead to personnel under-utilization. An optimistic approach in job-acceptance (having no or small time buffers) can lead to high outsourcing costs. This trade-off lies at the heart of the tactical model formulated below.

3. Model

On the tactical level, we assume that the allocation problem is carried out on full crews, which are categorized into classes based on the certificates of their members. Also, the jobs are categorized into job classes, according to their certificate requirements. Such structure makes the basic underlying model which we built-on close to the model presented in [6].

Let \mathcal{J} be the set of jobs, A be the set of job classes and \mathcal{C} be the set of crew classes. Each job j has an associated job class $a_j \in A$. For each crew class $c \in \mathcal{C}$, let $\mathcal{A}_c \subseteq A$ be the set of job classes that can be carried out by the crews in class c . For each job $j \in \mathcal{J}$, let $\mathcal{C}_j \subseteq \mathcal{C}$ be the set of crew classes which can carry out job j . Let p_j be the profit for processing job j . The number of crews in class c is bounded by M_c . The starting times s_j are fixed, but the finishing times are random variables $f_j(\xi)$ with known probability distribution. The problem can be formulated as two-stage problem with recourse:

$$\max \sum_{j \in \mathcal{J}} \sum_{c \in \mathcal{C}} p_j x_{jc} - \mathbb{E}_{\xi} \left[\sum_{c \in \mathcal{C}} q_c y_c(\xi) \right]$$

$$\begin{aligned}
 y_c(\xi) &\geq \sum_{j: a_j \in \mathcal{A}_c \wedge s_j \leq t < f_j(\xi)} x_{jc} - M_c, & \xi \in \Xi, t \in \mathcal{T}, c \in \mathcal{C}, \\
 \sum_{c \in \mathcal{C}} x_{jc} &\leq 1, & j \in \mathcal{J}, \\
 x_{jc} &\in \{0, 1\}, & c \in \mathcal{C}, j \in \mathcal{J}, \\
 x_{jc} &= 0, & a_j \notin \mathcal{A}_c, \\
 y_c(\xi) &\in \mathbb{N}, & c \in \mathcal{C},
 \end{aligned}$$

where x_{jc} indicates whether job j is assigned to a crew in crew class c , \mathcal{T} denotes the set of starting times³, $y_c(\xi)$ is the number of crews in class c which must be outsourced under scenario ξ , and q_c denotes the price for outsourcing. The objective is to maximize the profit minus the expected costs for outsourcing under the constraints that enough crews in each class are available to process all accepted jobs under each scenario and each job is processed by at most one crew class.

If $\bigcup_{c \in \mathcal{C}_1} \mathcal{A}_c \cap \bigcup_{c \in \mathcal{C}_2} \mathcal{A}_c = \emptyset$ for some $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$, $\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{C}$, the problem is separable and can be decomposed into two smaller problems. We assume that our problems are not separable.

4. Solutions

We have carried out experiments with simulated data based on the real-world data estimates. We have implemented the model in the GAMS modelling environment. For demonstration, we show below the data output for a moderate-size case consisting of 60 jobs, 10 crew classes, 60 time-periods and 15 scenarios. Each job can typically be processed by 2–3 crew classes and takes typically 3–8 days.

Table 1 shows the optimal values of the first-stage variables, and Table 2 shows the optimal values of the second-stage variables (outsourcing).

We have run the simulation repeatedly 30 times for various 15-scenario sets. The overall solution turned out to be rather stable, with variations of the objective function value less than 5%.

5. Conclusions and further research

The fixed interval scheduling problem with uncertain job durations and outsourcing, solved on tactical level, is highly relevant for practical use in demand-oriented logistics. We have carried out the logical analysis of the prob-

³It is not necessary to consider the finishing times.

Table 1: Optimal job acceptance and allocation

Job	Crew class C01–C10									
1	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	1	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
...										

Table 2: Optimal outsourcing of crews in crew classes

Crew class	Scenario 1–15														
C01	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C03	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1
C04	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
C05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C07	2	1	1	2	2	2	1	2	1	1	2	1	1	2	1
C08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

lem including a review of available academical results in this area. Based on our findings, we have formulated an original two-stage stochastic programming model and performed computational experiments based on quasi-real data. The results indicate that our approach is a viable one. Further work may be pointed in two directions. One direction is to provide more alternatives of the tactical model, including e.g. probability constraints, to model further advanced aspects of the real problem. Second direction can consist in detailing the operational part of the model – e.g. to provide a formulation where crew members can be considered individually.

Acknowledgements

The present work has been supported by the Molde University College's project NRF Power Up Project – WP3 and by European Regional Development Fund in the framework of the research project NETME Center under the Operational Program Research and Development for Innovation. Reg. r. CZ.1.05/2.1.00/01.0002, id code: ED0002/01/01.

References

- [1] Van den Bergh, J., Belien, J., De Bruecker, P., Demeulemeester, E., De Boeck, L.: Personnel scheduling: A literature review. *European Journal of Operational Research* 226, pp. 367–385, 2013.
- [2] Ernst, A. T., Jiang, H., Krishnamoorthy, M., Owens, B., Sier, D.: An Annotated Bibliography of Personnel Scheduling and Rostering. *Annals of Operations Research* 127, pp. 21–144, 2004. Kluwer Publishers.
- [3] Garey, M. R., Johnson, D. S.: *Computers and intractability: A guide to the theory of NP-completeness*. 1979. W.H. Freeman, San Francisco. ISBN 0-7167-1044-7.
- [4] Kolen, A. W. J., Lenstra, J. K., Papadimitriou, C. H., Spieksma, F. C. R.: Interval scheduling: A survey. *Naval Research Logistics* 54(5), pp. 530–543, 2007.
- [5] Kovalyov, M. Y., Ng, C. T., Cheng, T. C. E.: Fixed interval scheduling: Models, applications, computational complexity and algorithms. *European Journal of Operational Research* 178, pp. 331–342, 2007.
- [6] Kroon, L. G., Salomon, M., Van Wassenhove, L. N.: Exact and approximation algorithms for the operational fixed interval scheduling problem. *European Journal of Operational Research* 82, pp. 190–205, 1995.
- [7] Kroon, L. G., Salomon, M., Van Wassenhove, L. N.: Exact and approximation algorithms for the tactical fixed interval scheduling problem. *Operations Research* 45(4), pp. 624–638, 1997.
- [8] Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G., Shmoys, D. B.: Sequencing and scheduling: Algorithms and complexity. In Logistics of Production and Inventory, Vol. 4 of *Handbooks in Operations Research and Management Science*, pp. 445–522, 1993. Amsterdam: North-Holland.
- [9] Moehring, R. H., Radermacher, F. J., Weiss, G.: Stochastic scheduling problems I: General strategies. *ZOR – Zeitschrift fuer Operations Research* 28, pp. 193–260, 1984.
- [10] Nino-Mora, J.: Stochastic scheduling. In *Encyclopedia of Optimization*, C. A. Floudas, P. M. Pardalos (eds.), Vol. V, pp. 367–372, 2001. Kluwer.

POUŽITÍ LOGISTICKÉ REGRESE PRO MALÉ DATOVÉ VÝBĚRY

LOGISTIC REGRESSION APPLIED TO SMALL DATA SAMPLES

Josef Ditrich

Address: Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics in Prague, Náměstí W. Churchilla 4, 130 67 Prague 3

E-mail: pepa.ditrich@seznam.cz

Abstrakt: Logistickou regresi se doporučuje použít při analýzách rozsáhlejších datových vzorků. V případě výběrů malého rozsahu, nebo pokud některá vysvětlující proměnná velmi přesně predikuje některou z hodnot vysvětlované proměnné, parametry odhadnuté metodou maximální věrohodnosti mohou být vychýlené. K podobným problémům dochází i při analýzách kontingenčních tabulek, kdy v mnoha buňkách tabulky jsou nízké hodnoty sdružených četností. Problémy při odhadu parametrů modelu jsou způsobeny jevem známým jako separace (kompletní nebo částečná). Jedním ze způsobů, jak se s nedostatkem dat vypořádat, je použití exaktní logistické regrese. Alternativně lze využít některou z metod založenou na bayesovském přístupu. Tento článek se zaměřuje na použití logistické regrese pro malé výběry. Na simulovaných datech je porovnána rozdílnost a spolehlivost tří možných přístupů: metoda maximální věrohodnosti, kterou reprezentuje „klasická“ logistická regrese, exaktní logistická regrese a Firthova penalizační metoda maximální věrohodnosti zastupující bayesovský přístup. Výsledky ukazují, že v případě zjištění přítomnosti separace v datech nelze očekávat smysluplné výsledky při použití metod založených na „tradiční“ metodě maximální věrohodnosti.

Klíčová slova: malé datové výběry, logistická regrese, exaktní logistická regrese, Firthova penalizační metoda maximální věrohodnosti.

Abstract: Logistic regression is recommended for application to larger data sets. If the sample size is small or if any predictor is strongly associated with one of the possible outcomes, the estimated coefficients may be biased. A similar problem occurs in contingency tables when too many cells in the table have low counts. This phenomenon, known as separation (complete or partial separation) will cause problems when applied to the model. One approach to handle this type of problems is to employ exact logistic regression. Another option is to use a Bayesian approach. The paper focuses on the application

of logistic regression in the case of small data sets. For these purposes, three methods are compared on simulated data: the maximum likelihood method presented by logistic regression, exact logistic regression and Firth's penalized maximum likelihood method as a represent of a Bayesian approach. It has been identified that in the case of problem separation, the methods based on the 'traditional' maximum likelihood method cannot be used while the exact logistic regression and Firth's approach provides sensible outputs.

Keywords: small data samples, logistic regression, exact logistic regression, Firth's penalized maximum likelihood method.

1. Introduction

Logistic regression (LR) is used for its simplicity, accessibility and ease of interpretation of parameters. This method of statistical analysis is often used for developing unbiased models and determining the relationship among variables in data sets (Heinze and Schemper [11]).

However, in cases of small databases or rare events parameter estimates gained by the 'traditional' maximum likelihood method (MLM) are not reliable and even may not exist. With the decreasing sample size, chances of biased calculations increase as well as the probability that MLM used for parameter estimates will not converge to the maximum. This could be due to the presence of complete or partial separation. In extreme cases, estimates of parameters can reach plus or minus the point of infinity (Allison [4]). Therefore it is essential to test whether any type of separation is present in data, especially when binary outcome is modelled under conditions when the size of data set is small. The occurrence of a separation problem is typical for medical sciences (biostatistics), marketing researches or for certain portfolios in credit scoring (e.g., models for developers, investors or top affluent clients) (Heinze [12]).

The rest of this paper is organized as follows. The next section presents a brief overview of the separation problem. Section 3 explains how to handle the separation presence in data and provides an outline of usable methods. Section 4 contains a simulated case study and discusses analyses results. The final section presents conclusions.

2. Separation problem definition

When one of explanatory variable can perfectly classify observations into groups of the explained variable, the likelihood function has no maximum. Therefore, no finite values can be calculated for the vector of parameters

estimates $\widehat{\beta}$. This problem is referred to as monotone likelihood. Albert and Anderson [2] distinguish three different classes into which the underlying data can be classified – completely separated, partially (quasi-completely) separated and overlapping data. When one of the types of separation is present in data, only an infinite or a zero maximum likelihood (ML) estimate for the odds ratio is obtained.

2.1. Complete separation

Complete separation is present in a data sample if there is a vector of coefficients β such that when $\beta^T \mathbf{x}_i > 0$ then $y_i = 1$ and when $\beta^T \mathbf{x}_i < 0$ then $y_i = 0$. It means that a linear function of the explanatory variable X_i can perfectly predict values of the explained variable. An example of this situation is exhibited in Table 1. The example illustrates the model where there is only one independent variable X while dependent variable Y has value $y_i = 0$ whenever $x_i < 0$ and $y_i = 1$ whenever $x_i \geq 0$.

Table 1: Example of Complete Separation, Source: Allison [3]

i	1	2	3	4	5	6	7	8	9	10
x_i	-5	-4	-3	-2	-1	1	2	3	4	5
y_i	0	0	0	0	0	1	1	1	1	1

The observations in Table 1 can be summarised into the contingency table (see Table 2). Complete separation in the table (2×2) is made in case of joint frequencies $n_{11} = n_{22} = 0$ or $n_{12} = n_{21} = 0$.

Table 2: Example of Contingency Table in Case of Complete Separation

Condition	$y = 0$	$y = 1$
$x < 0$	5	0
$x \geq 0$	0	5

If complete separation exists within a sample, the ML estimate does not exist because the expression (ML estimator) $\widehat{\beta} = \log \frac{5 \times 5}{0 \times 0}$ is not defined. Even though the log likelihood is bounded by 0, no significant value for parameter β can be estimated (Allison [3]). The standard error will be infinite in size since the diagonal elements in the variance matrix are also infinite in size.

2.2. Partial separation

Partial separation is identified if there is a vector of coefficients β such that $y_i = 1$ when $\beta^T \mathbf{x}_i \geq 0$ and $y_i = 0$ when $\beta^T \mathbf{x}_i \leq 0$ and for at least one category of the explained variable Y the equality holds.

As an example of partial separation, let us have again only one independent variable X in the model. The dependent variable Y has value $y_i = 0$ whenever $x_i < 0$ and $y_i = 1$ whenever $x_i > 0$. However, there are cases for one value of X for which $y_i = 0$ and $y_i = 1$ can be observed.

Table 3: Example of Partial Separation, Source: Allison [3]

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	-5	-4	-3	-2	-1	0	0	1	2	3	4	5
y_i	0	0	0	0	0	0	1	1	1	1	1	1

By summarizing the observations in Table 3, we can get a two-way contingency table, as shown in Table 4. For example, partial separation in the table (2×2) is developed in the case when joint frequency $n_{21} = 0$ and other joint frequencies (diagonal and off-diagonal) are not equal to 0.

Table 4: Example of Contingency Table in Case of Partial Separation

Condition	$y = 0$	$y = 1$
$x \leq 0$	6	1
$x > 0$	0	5

The ML estimator is in the form $\hat{\beta} = \log \frac{6 \times 5}{0 \times 1}$ and is not also defined. Even though the log likelihood is bounded by -1 in this case, no finite value for parameter β can be estimated (Allison [3]). The standard error will be infinite in size.

Since, in practice, partial separation is more common than complete separation, the simulation study was performed on the partially separated data.

2.3. Overlap

The situation, when neither complete nor partial separation in data is present, is called overlap. In this case it is assumed that a unique and finite solution

exists for ML estimates. For the calculation of ML estimates, the Newton-Raphson algorithm can be used, for instance (Silvapulle [15]).

When constructing a logistic regression model, it is, especially in conditions of small samples or rare events, essential to test whether the separation is present in a data set. If so, it is necessary to choose the appropriate approach to handle the problem in order to interpret the estimated regression coefficients correctly. Some of them are described in the next section.

3. Separation problem solution

The problem of separation was firstly documented by Day and Kerridge [8]. Since that time, many methods of detecting the presence of complete or partial separation have been proposed. A linear programming method indicating non-overlapping data was introduced by Albert and Anderson [2]. It was further improved by Santner and Duffy [14] with the goal to distinguish all three states described above. Christmann and Rousseeuw [5] introduced a method for assessing the ‘degree’ of data overlapping.

Methods used to deal with separation can be broadly divided into two groups. The first group includes approaches which suppose that the separation is a data problem. An easy and often used method in practice is to exclude the problematic variable causing the separation from the model development. However, this method cannot be recommended because omitting the ‘strong’ variable can cause specification bias, which can lead to biased parameter estimates. Another method proposed by Clogg et al. [6] is to add artificial data across various patterns to fill the gaps in order to change the covariates originally included into the model. However, in Heinze and Schemper [11], it is shown that this method is not appropriate. In case of a nominal variable, merging categories can be an option. Yet, it should be kept in mind that it will not be impossible to allocate the observations to the original categories.

The second group represents the methods of adjusting the likelihood function. These include exact logistic regression (ELR), Firth’s penalized maximum likelihood method (FPMLM) and hidden logistic regression (HLR). HLR, the concept of which was introduced by Rousseeuw and Christmann [13] in 2003, will not be discussed further in the paper.

3.1. Exact logistic regression

One way to deal with separation is to use exact logistic regression (also known as exact conditional inference), as proposed by Cox and Snell [7]. It is an approach based on the widespread methodology of Fisher’s exact test for

(2 × 2) contingency tables. Unknown parameter is estimated by maximizing conditional likelihood function.

Let us have a vector of parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ and a vector of sufficient statistics $\mathbf{t} = (t_0, t_1, \dots, t_p)$ for $\boldsymbol{\beta}$. The sufficient statistic t_j for parameter β_j is $t_j = \sum_{i=1}^n y_i x_{ij}$. The t_j value contains all the information about β_j for observed data. The conditional likelihood function l_c of one specific parameter β_j can be expressed by using the unconditional likelihood function l and eliminating the effect of the rest of the parameters $\beta_0, \beta_1, \dots, \beta_{j-1}$ as (Agresti [1]):

$$l_c(\beta_j) = \frac{c(t_0, t_1, \dots, t_j) \cdot e^{\beta_j t_j}}{\sum_u c(t_0, t_1, \dots, t_{j-1}, u) \cdot e^{\beta_j u}}, \quad j = 0, 1, \dots, p, \quad (1)$$

where $c(t_0, t_1, \dots, t_j)$ denotes the number of distinct binary sequences (number of possible combinations of 0 and 1) for alternative dependent variable that give the specific values t_0, t_1, \dots, t_j and summation term in the denominator is over all possible values of u for which $c(t_0, t_1, \dots, t_{j-1}, u) \geq 1$.

To use ELR has number of limitations. The estimations cannot be calculated, if there is no other combination of data vector except of values currently observed (Heinze and Schemper [11]). According to Allison [4], ELR cannot be used, if there is more than one variable causing separation in a data sample. When using a combination of both categorical and continuous explanatory variables, unreliable estimates can be obtained (Zorn [16]).

3.2. Firth's penalized maximum likelihood method

The second method to solve non-overlapping data is by penalizing the likelihood function, as proposed by Firth [9]. This enhances the ability of the penalized function to converge to its maximum. In Heinze and Schemper [11], the method is extended by the calculation of confidence intervals. The method was further developed by Gao and Shen [10] for those cases when multicollinearity is present in data.

FPMLM develops an idea to remove biased computation in phase of parameters estimation. This method adjusts original likelihood function $l(\boldsymbol{\beta})$ by multiplying it by the square root of determinant of Fisher's information matrix $\mathbf{I}(\boldsymbol{\beta})$, which is given by $\mathbf{I}(\boldsymbol{\beta}) = \phi^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X}$, where for a binomial distribution $\phi = 1$ and \mathbf{W} is a $(n \times n)$ matrix where $\mathbf{W} = \text{diag}(\pi_i(1 - \pi_i))$. The adjusted likelihood function $l^*(\boldsymbol{\beta})$ (and its logarithmized form for calculation purposes) is formulated as:

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) \sqrt{|\mathbf{I}(\boldsymbol{\beta})|}. \quad (2)$$

$$\log l^*(\boldsymbol{\beta}) = \log l(\boldsymbol{\beta}) + \frac{1}{2} \log |\mathbf{I}(\boldsymbol{\beta})|. \quad (3)$$

To find the maximum of penalized likelihood function for one specific parameter β_j , the first derivative $U^*(\beta_j)$ is set equal to 0:

$$U^*(\beta_j) = \frac{\partial \log l^*(\boldsymbol{\beta})}{\partial \beta_j} = U(\beta_j) + \frac{1}{2} \left\{ (\mathbf{I}(\boldsymbol{\beta}))^{-1} \left(\frac{\partial \mathbf{I}(\boldsymbol{\beta})}{\partial \beta_j} \right) \right\} = 0, \\ j = 0, 1, \dots, p, \quad (4)$$

where $U(\beta_j) = \frac{\partial \log l(\boldsymbol{\beta})}{\partial \beta_j}$. Depending on the above derivative, the parameter estimate is attained by the iterative process:

$$\hat{\beta}_{j+1} = \hat{\beta}_j - (\mathbf{I}(\hat{\beta}_j))^{-1} U^*(\hat{\beta}_j). \quad (5)$$

Tests based on likelihood ratio are recommended instead of Wald's tests especially when separation is present (Firth [9]).

Heinze [12] highlights advantages of FPMLM over ELR. The major difference between the two could be identified in that the former could be used for different amounts and character of variables. The method is also recognized to be used on computers to a minimum extent for inadequate software performance.

4. Simulation study

To illustrate the consequences of partially separated data, the following simulation study was performed. It was created on a data set with 20 observations, while it was considered that it is necessary to build a model where only one dichotomous explanatory variable X will be included into the model to predict two values of a dependent variable Y . Values of variable X were generated by using a random numbers generator where probability that $x = 0$ was set to 0.4 and $x = 1$ equalled to 0.6. To get values of Y , logistic regression was used with the regression coefficients $\beta_0 = -15$ and $\beta_1 = 16$. To ensure that partial separation will be present, auxiliary variable A was created and filled with random numbers from uniform distribution, while the following conditions were applied: if $\frac{e^{\beta_0 + \beta_1 \cdot x_i}}{1 + e^{\beta_0 + \beta_1 \cdot x_i}} > a_i$ then $y_i = 1$, otherwise $y_i = 0$. Subsequently, all three models were build and the results were collected. This process was repeated 1,000 times and the received results were averaged.

As a result of the previously described process, the statistics for observed joint frequencies are stated in Table 5. Here, \bar{p} stands for mean, \bar{p}_{ci} is 95% confidence interval for a mean, p_5 and p_{95} are 5th and 95th percentiles and p_{50} represents a median.

Table 5: Joint Frequencies – Simulation Study in Case of Partial Separation

	$y = 0$	$y = 1$
$x = 0$	$\bar{p} = 7.95, \quad \bar{p}_{ci}(7.82, 8.09)$ $p_5 = 5, \quad p_{95} = 12$ $p_{50} = 8$	0
$x = 1$	$\bar{p} = 3.41, \quad \bar{p}_{ci}(3.31, 3.51)$ $p_5 = 1, \quad p_{95} = 6$ $p_{50} = 3$	$\bar{p} = 8.76, \quad \bar{p}_{ci}(8.62, 8.90)$ $p_5 = 5, \quad p_{95} = 13$ $p_{50} = 9$

The following Table 6 shows estimations of parameters gained by all three methods. For these purposes, LOGISTIC procedure in SAS 9.4 (with appropriate option) was employed.

Table 6: Results of Methods, Average Values (*Median Unbiased Estimate)

Method	β_0	[p -value]	(95% conf. int.)
LR	-12.21	[0.9373]	(-342.26; 317.83)
ELR	-2.35*	[0.0231]	($-\infty$; -0.72)
FPMLM	-3.05	[0.0493]	(-5.84; -0.26)

Method	β_1	[p -value]	(95% conf. int.)
LR	13.51	[0.9301]	(-317.58; 344.60)
ELR	3.03*	[0.0171]	(1.16; $+\infty$)
FPMLM	3.78	[0.0386]	(0.42; 7.13)

Outputs gained by MLM do not provide a proper interpretation. Parameter estimates for LR are high and determined in absolute values in comparison to alternative approaches. Parameters are not statistically significant according to p -values of Wald test. Therefore, it can be determined that the usage of LR in the case of existence of separation is not suitable.

On the other hand, based on the observation made on FPMLM and ELR, it can be identified that both parameter estimates are significant. Interval estimations of both parameters do not contain value 0. That supports the conclusion that values of categorical explanatory variable X affect binary

outcome Y . This suggests that the computations that are conducted with logistic regression using various sample sizes needed to be estimated based on the proper mode of iteration that are to be used.

5. Conclusion

In the paper, the problem of complete and partial separation in data sets was discussed. Further, a brief introduction of how to solve the separation problem was given and two suitable statistical methods mentioned. Both methods were tested against the logistic regression. For this purpose, a small data set of twenty observations was repeatedly generated and the presence of the partial separation was ensured. Data samples purposely contained only one categorical covariate that had to be used to predict a binary outcome.

It has been explained in the paper that if separation is present in a data sample, then it is not suitable to use approaches based on the ‘traditional’ maximum likelihood method as logistic regression is. The review of the results shows that alternative approaches should be used when dealing with data sets where separation is present. This even points to the fact that analysts could reveal a data problem before any method is performed and could choose an appropriate one from among the available options of computation.

Many different solutions for completely and partially separated data exist in practice. Each method has its own pros and cons, which should be investigated prior to their application. While ‘classical’ logistic regression is best used on large data samples, the other two examined methods give better results of small data samples. Exact logistic regression performs well with nominal explanatory variables. According to referred literature, Firth’s approach has a more universal usage. It can be applied in both cases – when continuous as well as nominal covariates need to be included.

References

- [1] Agresti, A.: *Categorical Data Analysis*. Third edition. John Wiley & Sons, New York, 2012, ISBN 9780470463635.
- [2] Albert, A., Anderson, J.: On the Existence of Maximum Likelihood Estimates in Logistic Regression. *Biometrika* 71, pp. 1–10, 1984.
- [3] Allison, P.: Convergence Problems in Logistic Regression. In Altman, M., Gill J., McDonald, M. P.: *Numerical Issues in Statistical Computing for the Social Scientist*, John Wiley & Sons, pp. 219–233, 2004.
- [4] Allison, P. D.: *Convergence Failures in Logistic Regression*. SAS Global Forum 2008, paper 360, 2008.

- [5] Christmann, A., Rousseeuw, P. J.: Measuring Overlap in Logistic Regression. *Computational Statistics and Data Analysis* 37, pp. 65–75, 2001.
- [6] Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B., Weidman, L.: Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples using Bayesian Logistic Regression. *Journal of the American Statistical Association* 86, pp. 68–78, 1991.
- [7] Cox, D. R., Snell, E. J.: *Analysis of Binary Data*. Second edition. CRC Press, 1989, ISBN 9780412306204.
- [8] Day, N. E., Kerridge, D. F.: A General Maximum Likelihood Discriminant. *Biometrics*, pp. 313–323, 1967.
- [9] Firth, D.: Bias Reduction of Maximum Likelihood Estimates. *Biometrika* 80, pp. 27–38, 1993.
- [10] Gao, S., Shen, J.: Asymptotic Properties of a Double Penalized Maximum Likelihood Estimator in Logistic Regression. *Statistics & Probability Letters* 77, pp. 925–930, 2007.
- [11] Heinze, G., Schemper, M.: A Solution to the Problem of Separation in Logistic Regression. *Statistics in Medicine* 21, pp. 2409–2419, 2002.
- [12] Heinze, G.: A Comparative Investigation of Methods for Logistic Regression with Separated or Nearly Separated Data. *Statistics in Medicine* 25, pp. 4216–4226, 2006.
- [13] Rousseeuw, P. J., Christmann, A.: Robustness Against Separation and Outliers in Logistic Regression. *Computational Statistics and Data Analysis* 43, pp. 315–332, 2007.
- [14] Santner, T. J., Duffy, E. D.: A Note on A. Albert and J. A. Anderson's Conditions for the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* 73, pp. 755–758, 1986.
- [15] Silvapulle, M. J.: On the Existence of Maximum Likelihood Estimators for the Binomial Response Models. *Journal of the Royal Statistical Society Series B*, pp. 310–313, 1981.
- [16] Zorn, C.: A Solution to Separation in Binary Response Models. *Political Analysis* 13, pp. 157–170, 2005.

STATISTICKÁ GRAMOTNOST V RÁMCOVÝCH VZDĚLÁVACÍCH PROGRAMECH

STATISTICAL LITERACY FRAMEWORK EDUCATION PROGRAMMES

Aneta Hybšová

E-mail: anetahybsova@gmail.com

Abstrakt: Význam statistické gramotnosti roste v souvislosti s rozvojem vědy a výzkumu a především s rozvojem ICT technologií. Úroveň statistické gramotnosti se však rozvíjí velmi pomalu. Cílem tohoto příspěvku je zjistit, do jaké míry se objevuje statistika v rámcových vzdělávacích programech základních škol a gymnázií. Dle zjištění z analýzy rámcových vzdělávacích programů si žáci základních škol a gymnázií osvojí základní pojmy popisné statistiky a celou řadu klíčových kompetencí, což je vhodný výchozí stav pro vysokoškolské studium základních statistických kurzů.

Klíčová slova: statistická gramotnost, rámcový vzdělávací program, matematika a její aplikace, základní škola, ZŠ, střední škola, SŠ.

Abstract: The importance of statistical literacy is increasing in connection with development of research and ICT technologies. But the level of statistical literacy development is lower. The aim of this paper is to find out the rate of implementation of statistics into framework education programme and how statistical literate students are admitted to university.

Keywords: statistical literacy, framework education programme, mathematics and its application, basic school, high school.

1. Úvod – význam statistické gramotnosti

Význam statistické gramotnosti, tedy schopnosti vytvořit, porozumět a kriticky zhodnotit statistické výsledky (definice dle [11]), je v dnešní informační době nesporný. Statistika dává odpověď na otázku, jak vyhodnotit a chápat data získaná empirickým výzkumem. Vzhledem k tomu, že statistické výzkumy jsou dnes běžně používány v praxi a výsledky prezentovány společnosti, je i výuka statistiky čím dál více implementována nikoliv pouze do matematicky orientovaných vysokoškolských studijních oborů, ale vyučuje se v modifikované formě i v oborech nematematicky zaměřených, viz [4] a [10]. Statistické uvažování je po nás vyžadováno každý den [2, str. 21] a i vzhledem k tomu, že média často používají čísla a různé statistiky jako argumenty a předkládají je občanům, je důležité, aby občané těmito informacím rozuměli

a kriticky je analyzovali [5]. Právě porozumění těmto informacím a kritické myšlení může lidem pomoci správně se rozhodovat. Otázkou však je, kde se občan naučí takovým schopnostem.

Statistická gramotnost se stejně jako většina gramotností rozvíjí po celou dobu života. Právě na základní a střední škole, kdy se myšlení žáků formuje, je významné učit žáky pracovat s informacemi a věrohodností jejich zdroje. Nasnadě je tedy otázka, co by měli studenti nastupující na vysoké školy umět a jaké kompetence v oblasti statistické gramotnosti by měli mít. K nalezení odpovědi na tuto otázku byly analyzovány rámcové vzdělávací programy základních škol (dále jen RVP ZŠ) a gymnázií (dále jen RVP G).

2. Metodika

Za účelem zjištění úrovně statistické gramotnosti studentů nastupujících na vysoké školy byly prostudovány RVP ZŠ a RVP G. Gymnázia byla vybrána za zástupce středních škol, neboť lze předpokládat, že největší počet uchazečů o studium na vysoké škole pochází právě z gymnázií.

Práce je fakticky rozdělena do dvou částí. Nejprve byla provedena analýza vzdělávací oblasti *Matematika a její aplikace*, neboť je statistika v současné době vyučována v rámci předmětu matematika, a tudíž lze předpokládat, že nejvíce znalostí pojmů a statistického myšlení si žáci osvojí právě v rámci této vzdělávací oblasti. Následně jsou analyzovány i další vzdělávací oblasti, které, ačkoliv nemají se statistikou zdánlivě mnoho společného, obsahují nejrůznější schopnosti a dovednosti, které velmi úzce souvisí se statistickou gramotností.

Zkoumání vzdělávací oblasti *Matematika a její aplikace* bylo provedeno jak v RVP ZŠ, tak i v RVP G. V rámci analýzy byly vyhledány statistické pojmy a dovednosti, které by si žáci měli v průběhu studia osvojit. Tyto pojmy a dovednosti byly zaznamenány do shrnující Tabulky 1. V dalších vzdělávacích oblastech (tedy mimo vzdělávací oblast *Matematika a její aplikace*) byly systematicky vyhledány následující pojmy úzce související se statistickou gramotností:

- | | |
|------------------------|----------------------|
| • Statistika | • Model |
| • Data | • Analytické myšlení |
| • Závislost | • Kritické myšlení |
| • Tabulka | • Informace |
| • Graf | • Analýza |
| • Tabulkový kalkulátor | • Empirické metody |
| • Tabulkový procesor | • Výzkum |

Tyto pojmy (včetně všech jejich slovních tvarů) a jejich souvislost se statistickou gramotností jsou popsány v kapitole Přesahy statistické gramotnosti do dalších vzdělávacích oblastí a v kapitole Přesahy statistické gramotnosti do průřezových témat na základních školách a gymnáziích.

3. Statistika ve vzdělávací oblasti Matematika a její aplikace

V rámci analýzy byla tedy podrobena zkoumání nejprve vzdělávací oblast **Matematika a její aplikace**. Zkoumána byla nejprve tato oblast v RVP ZŠ a následně v RVP G.

3.1. Statistika ve vzdělávací oblasti Matematika a její aplikace na základních školách

Vzdělávací oblast *Matematika a její aplikace* je součástí předmětu matematika a v RVP ZŠ je popsána jako oblast založená na aktivních činnostech, které jsou typické pro práci s matematickými objekty a pro užití matematiky v reálných situacích. Tato oblast umožňuje získávat matematickou gramotnost, která úzce souvisí s gramotností statistickou. Matematika i statistika poskytuje vědomosti a dovednosti potřebné v běžném životě a vytváří předpoklady pro další úspěšné studium.

Vzdělávací obsah vzdělávací oblasti *Matematika a její aplikace* je rozdělen na čtyři tematické okruhy: *Čísla a početní operace* probíraném na prvním stupni, tematický okruh *Číslo a proměnná* probíraném na druhém stupni, *Závislosti, vztahy a práce s daty* a *Geometrie v rovině a v prostoru*.

Pro rozvoj statistické gramotnosti je nejvýznamnější tematický okruh *Závislosti, vztahy a práce s daty*, kde žáci rozpoznávají závislosti, které jsou projevem běžných jevů v reálném světě. Tyto závislosti se žáci učí analyzovat pomocí tabulek, diagramů a grafů. V jednoduchých případech tyto závislosti konstruují pomocí matematického modelu – ideálně pomocí vhodného softwaru nebo grafického kalkulátoru. Mezi očekávané výstupy tohoto okruhu patří schopnost žáka vyhledat, zpracovat a vyhodnotit data, porovnat soubory dat, určit vztah přímé anebo nepřímé úměrnosti, vyjádřit funkční vztah tabulkou, rovnicí, grafem a matematizovat jednoduché reálné situace s využitím funkčních vztahů.

Učivo tohoto okruhu je shrnuto do dvou podoblastí – závislosti a data (příklady závislostí z praktického života a jejich vlastnosti, nákresy, schémata, diagramy, grafy, tabulky, četnost znaku, aritmetický průměr) a funkce

(pravoúhlá soustava souřadnic, přímá úměrnost, nepřímá úměrnost, lineární funkce) [8, str. 32]. Mezi cíle tohoto tematického okruhu patří například

- rozvíjení logického myšlení ke kritickému usuzování, srozumitelné a věcné argumentaci prostřednictvím řešení matematických problémů,
- vnímání složitosti reálného světa a jeho porozumění,
- provádění rozboru problému a plánu řešení, odhadování výsledků, volba správného postupu k vyřešení problému a vyhodnocování správnosti výsledku vzhledem k podmínkám úlohy nebo problému,
- rozvíjení spolupráce při řešení problémových a aplikovaných úloh vyjadřujících situace z běžného života a následně k využití získaného řešení v praxi; k poznávání možností matematiky a skutečnosti, že k výsledku lze dospět různými způsoby,
- rozvíjení důvěry ve vlastní schopnosti a možnosti při řešení úloh, výchova k soustavné sebekontrolě při každém kroku postupu řešení, k rozvíjení systematickosti, vytrvalosti a přesnosti, k vytváření dovednosti vyslovovat hypotézy na základě zkušenosti nebo pokusu a k jejich ověřování nebo vyvracení pomocí protipříkladů. [8]

Statistika se v této vzdělávací oblasti může prolínat s ICT předměty, neboť různé závislosti lze žákům prezentovat pomocí software nebo grafických kalkulátorů. Právě tím, že se žáci učí využívat prostředky výpočetní techniky (především kalkulátory, vhodný počítačový software, určité typy výukových programů) a používat některé další pomůcky, je umožněn přístup k matematice i žákům, kteří mají nedostatky v numerickém počítání a v rýsovacích technikách. Zdokonalují se rovněž v samostatné a kritické práci se zdroji informací.

3.2. Statistika ve vzdělávací oblasti Matematika a její aplikace na gymnáziích

Dalším krokem po analýze RVP ZŠ byla analýza RVP G. Výuka matematiky na gymnáziu rozvíjí a prohlubuje pochopení kvantitativních a prostorových vztahů reálného světa, utváří kvantitativní gramotnost žáků [1]. V RVP G je matematika obsažena ve vzdělávací oblasti *Matematika a její aplikace*, která je součástí předmětu Matematika. Tato vzdělávací oblast obsahuje pět tematických okruhů *Argumentace a ověřování*, *Číslo a proměnná*, *Práce s daty a kombinatorika*, *pravděpodobnost*, *Závislost a funkční vztahy*, *Geometrie*.

Z pohledu statistické gramotnosti je nejzajímavější okruh *Práce s daty a kombinatorika, pravděpodobnost*, v rámci něhož se žáci učí řešit reálné

problémy s kombinatorickým podtextem, pomocí kombinatorických postupů vypočítat pravděpodobnosti náhodných jevů, diskutovat a kriticky hodnotit statistické informace a statistická sdělení, volit a využívat vhodné statistické metody k analýze a zpracování dat (včetně využití výpočetní techniky). V neposlední řadě se pak žáci učí graficky reprezentovat soubor dat a interpretovat tabulky, diagramy a grafy. Žák se také naučí porovnat dva různé soubory dat vzhledem k jejich odlišným charakteristikám. V rámci tohoto tematického okruhu si žák osvojí pojmy vážený aritmetický průměr, modus, medián, percentil, kvartil, směrodatná a mezikvartilová odchylka.

Dalším zajímavým tematickým okruhem je *Argumentace a ověřování*. Mezi očekávané výstupy patří mimo jiné schopnost žáka vytvořit hypotézy, zdůvodnit jejich pravdivost či nepravdivost a vyvrátit nesprávná tvrzení. Dalším očekávaným výstupem je schopnost žáka zdůvodnit svůj postup a ověřit správnost řešení problému.

3.3. Shrnutí osvojených statistických pojmů na základních školách a gymnáziích

Po prozkoumání obou dokumentů – RVP ZŠ a RVP G, byly pojmy a dovednosti, které by si žáci měli osvojit v průběhu studia na základní škole a gymnáziu, shrnuty do zřehledňující Tabulky 1.

Tabulka 1: Shrnutí osvojených pojmů a dovedností ve vzdělávací oblasti Matematika a její aplikace na základních školách a gymnáziích

Matematika a její aplikace		
Osvojené	Základní školy	Gymnázia
Pojmy	<ul style="list-style-type: none"> ● Graf ● Diagram ● Tabulka ● Statistický soubor ● Kvantitativní znak ● Kvalitativní znak ● Četnost ● Modus ● Medián ● Aritmetický průměr 	<ul style="list-style-type: none"> ● Náhodný jev ● Pravděpodobnost ● Statistický soubor ● Charakteristiky statistického souboru ● Aritmetický průměr ● Vážený aritmetický průměr ● Modus ● Medián ● Percentil ● Kvartil ● Směrodatná odchylka ● Mezikvartilová odchylka

Dovednosti	<ul style="list-style-type: none"> ● Přečte z grafu požadované údaje. ● Orientuje se v grafech. ● Provádí statistická šetření a zapisuje je pomocí tabulky nebo diagramu. ● Vyhledává, vyhodnocuje, zpracovává data. ● Porovnává soubory dat. ● Pracuje s intervaly a časovou osou. ● Samostatně vyhledává data v literatuře, denním tisku a na internetu a kriticky hodnotí jejich reálnost. 	<ul style="list-style-type: none"> ● Diskutuje a kriticky zhodnotí statistické informace a daná statistická sdělení. ● Volí a užívá vhodné statistické metody k analýze a zpracování dat (využívá výpočetní techniku). ● Reprezentuje graficky soubory dat, čte a interpretuje tabulky, diagramy a grafy, rozlišuje rozdíly v zobrazení obdobných souborů vzhledem k jejich odlišným charakteristikám.
------------	--	---

4. Přesahy statistické gramotnosti do dalších vzdělávacích oblastí

4.1. Přesahy do vzdělávacích oblastí na základní škole

Po analýze vzdělávacích oblastí *Matematika a její aplikace* na základních školách a gymnáziích, byly podrobně prozkoumány i další vzdělávací oblasti. Mimo vzdělávací oblast *Matematika a její aplikace* se statistika objevuje i ve vzdělávací oblasti oblast **Informační a komunikační technologie**, která, jak se píše v [1, str. 34]: „... umožňuje všem žákům dosáhnout základní úrovně informační gramotnosti – získat elementární dovednosti v ovládnutí výpočetní techniky a moderních informačních technologií, orientovat se ve světě informací, tvořivě pracovat s informacemi a využívat je při dalším vzdělávání i v praktickém životě.“ Právě práce s informacemi a využívání výpočetní techniky spojuje tuto vzdělávací oblast se statistikou, neboť právě zpracovávání dat ideálně pomocí tabulkových kalkulátorů je v dnešní době velice nezbytná dovednost. Vzdělávací obsah oblasti *Informační a komunikační technologie* je rozdělen do dvou tematických okruhů – *Vyhledávání informací a komunikace* a *Zpracování a využití informací*. Statistika se prolíná oběma oblastmi, v prvním případě souvisí především s vyhledáváním dat a jejich relevancí.

Žáci se tedy učí pracovat s různými databázemi, ročenkami, statistickými přehledy, a dále hledat nejrůznější informace na internetu a kriticky o nich smýšlet. V druhém okruhu se statistika především uplatňuje v rámci práce s tabulkovými procesory, zpracováním získaných dat a tvorbou grafických výstupů.

Mezi očekávané výstupy prvního tematického okruhu *Vyhledávání informací a komunikace* patří schopnost žáka ověřit věrohodnost informací a informačních zdrojů, posoudit jejich závažnost a vzájemnou návaznost. Učivo pak zahrnuje nejen znalosti a dovednosti v oblasti vývojových trendů informačních technologií, ale i znalost hodnoty a relevantnosti informací a informačních zdrojů včetně metod a nástrojů jejich ověřování.

Mezi očekávané výstupy druhého tematického okruhu *Zpracování a využití informací* patří schopnost žáka ovládat práci s tabulkovými editory a využívat vhodných aplikací, používat informace z různých informačních zdrojů a vyhodnocovat jednoduché vztahy mezi údaji, zpracovat a prezentovat na uživatelské úrovni informace v textové, grafické a multimediální formě. Učivo pak zahrnuje mimo jiné tabulkové editory, vytváření tabulek, porovnávání dat, jednoduché vzorce a prezentaci informací skrze webové stránky, prezentační programy, multimédia [1, str. 32].

Dovednosti získané v této vzdělávací oblasti umožňují žákům aplikovat výpočetní techniku ve všech vzdělávacích oblastech celého základního vzdělávání. Právě tato aplikační rovina přesahuje rámec vzdělávací oblasti *Informační a komunikační technologie* a stává se součástí všech vzdělávacích oblastí základního vzdělávání. Cílem této vzdělávací oblasti je rozvoj a podpora klíčových kompetencí směřující žáka k seznámení s moderními informačními a komunikačními technologiemi a jejich využíváním, porozumění toku informací až po jejich zpracování a využití v praxi.

Žák se učí vytvářet algoritmus určité činnosti, využívat a aplikovat výukový software k efektivnímu učení a správné organizaci práce. Učí se poznávat nejen pozitivní, ale také negativní vlivy internetu a respekt k ochraně osobních dat a údajů, ale také vztah k ochraně práv duševního vlastnictví programového vybavení. Tyto kompetence úzce souvisí i s kritickým smýšlením o informacích a hledáním důvěryhodných zdrojů.

Třetí vzdělávací oblastí, se kterou se statistika může prolínat, je vzdělávací oblast **Člověk a příroda**, která je společná pro fyziku, chemii, přírodopis a zeměpis. V této oblasti se utváří klíčové kompetence tím, že je žák především veden k:

- zkoumání přírodních faktů a jejich souvislostí s využitím různých empirických metod poznávání (pozorování, měření, experiment) i různých metod racionálního uvažování,
- potřebě klást si otázky o průběhu a příčinách různých přírodních procesů, které mají vliv i na ochranu zdraví, životů, životního prostředí a majetku, správně tyto otázky formulovat a hledat na ně adekvátní odpovědi,
- způsobu myšlení, které vyžaduje ověřování vyslovovaných domněnek o přírodních faktech více nezávislými způsoby,
- posuzování důležitosti, spolehlivosti a správnosti získaných přírodovědných dat pro potvrzení nebo vyvrácení vyslovovaných hypotéz či závěrů.

Právě zde lze s žáky využít mezioborových vazeb s matematikou a využít znalostí statistiky. V přírodovědných předmětech žák může porovnávat získaná data a vytvářet grafy znázorňující zjištěná fakta. Například v předmětu zeměpis je součástí osvojených dovedností i schopnost organizovat a přiměřeně hodnotit geografické informace a zdroje dat z dostupných kartografických produktů a elaborátů, z grafů, diagramů, statistických a dalších informačních zdrojů.

Statistika a statistická gramotnost přesahuje i do dalších vzdělávacích oblastí. Například ve vzdělávací oblasti **Člověk a jeho svět** je souvislost se statistikou a statistickou gramotností evidentní. Kompetence k užívání matematického jazyka a příslušných pomůcek či nástrojů se v rámci této vzdělávací oblasti rozvíjí např. při práci s různými reprezentacemi dat, při používání náčrtků, plánek a map.

4.2. Přesahy do vzdělávacích oblastí na gymnáziích

Statistika se objevuje mimo vzdělávací oblast *Matematika a její aplikace* i v RVP G. Příkladem může být vzdělávací oblast **Člověk a příroda** zahrnující fyziku, chemii, biologii, geografii a geologii. Tyto přírodní disciplíny jsou si velmi blízké i v metodách a prostředcích, které uplatňují ve své výzkumné činnosti. Používají jak metody empirické (pozorování, měření a experimenty), tak i prostředky teoretické (pojmy, hypotézy, modely a teorie).

V RVP pro gymnázia se píše, že „Základní prioritou každé oblasti přírodovědného poznávání je odkrývat metodami vědeckého výzkumu zákonitosti, jimiž se řídí přírodní procesy. Odkrývání přírodních zákonitostí je hodnotné jednak samo o sobě, neboť naplňuje přirozenou lidskou zvědavost poznat a porozumět tomu, co se odehrává pod povrchem smyslově pozorovatelných, často

zdánlivě nesouvisejících jevů, a jednak člověku umožňuje ovládnout různé přírodní objekty a procesy tak, aby je mohl využívat pro další výzkum i pro rozmanité praktické účely.“ [1]. Z této definice vyplývá, že právě pomocí metod vědeckého výzkumu žáci s pomocí učitele odkrývají zákonitosti přírodních procesů. Žáci si v rámci této vzdělávací oblasti postupně osvojují vybrané empirické i teoretické metody přírodovědného výzkumu a aktivně je spolu s přírodovědnými poznatky ve výuce využívají. Žáci jsou vedeni k tomu, aby si uvědomili význam objektivitu a pravdivosti poznání. Těch lze ale dosahovat pouze pomocí nezávislé kontroly způsobu získávání dat a ověřování hypotéz. Gymnaziální vzdělávání v přírodovědné oblasti též musí vést ke kritickému smýšlení o problémech i pravdivosti předložených přírodovědných informací. Toho lze docílit tak, že žáci debatují o způsobech získávání dat a metodách ověřování hypotéz. Žákům je zapotřebí ukázat negativní důsledky zkreslování dat a seznámit je s morálním aspektem výzkumu.

Cílem vzdělávací oblasti je rozvíjet klíčové kompetence žáků. Z hlediska statistické gramotnosti obsahuje tato vzdělávací oblast tyto schopnosti:

- formulace přírodovědného problému,
- provádění soustavných a objektivních pozorování, měření a experimentů (především laboratorního rázu) podle vlastního či týmového plánu nebo projektu, k zpracování a interpretaci získaných dat a hledání souvislostí mezi nimi,
- tvorba modelu přírodního objektu či procesu umožňujícího pro daný poznávací účel vhodně reprezentovat jejich podstatné rysy či zákonitosti,
- používání adekvátních matematických a grafických prostředků k vyjadřování přírodovědných vztahů a zákonů,
- využívání prostředků moderních technologií v průběhu přírodovědné poznávací činnosti.

Další vzdělávací oblastí související s rozvojem statistické gramotnosti je vzdělávací oblast **Člověk a společnost** zahrnující občanský a společenskovední základ a dějepis. Tato oblast „... přispívá k utváření historického vědomí, k uchování kontinuity tradičních hodnot naší civilizace a k občanskému vzdělávání mládeže. Posiluje respekt k základním principům demokracie a připravuje žáky na odpovědný občanský život v demokratické společnosti...“ [1].

Vzdělávání v této vzdělávací oblasti směřuje k utváření a rozvíjení klíčových kompetencí tím, že vede žáka k utváření realistického pohledu na skutečnost a k orientaci ve společenských jevech a procesech tvořících rámeček

každodenního života a dále k osvojování demokratických principů v mezilidské komunikaci, k rozvíjení schopnosti diskutovat o veřejných záležitostech, rozpoznávat manipulativní strategie, zaujímat vlastní stanoviska a kritické postoje ke společenským a společenskovedním záležitostem, věcně (nepředpojatě) argumentovat.

Ve vzdělávací oblasti **Člověk a svět práce** je u žáka rozvíjena schopnost analyzovat působení médií v ekonomickém světě a využít aktuální mediální informace při analýze české i světové ekonomiky. Právě tato schopnost je klíčová právě pro kritické smýšlení o médiích publikovaných statistikách a jejich vzniku. Občanská statistická gramotnost je nezbytnou součástí statistické gramotnosti a umožňuje se plně orientovat v současném světě [6].

Poslední vzdělávací oblastí související se statistickou gramotností je vzdělávací oblast **Informatika a informační a komunikační technologie**, která navazuje na oblast ICT v základním vzdělávání zaměřenou na zvládnutí základní úrovně informační gramotnosti [1]. Vzdělávání v této vzdělávací oblasti směřuje k utváření a rozvíjení klíčových kompetencí tím, že vede žáka k porozumění zásadám ovládnutí a věcným souvislostem jednotlivých skupin aplikačního programového vybavení a k vhodnému uplatňování jejich nástrojů, metod a vazeb k efektivnímu řešení úloh a dále jej vede k využívání prostředků ICT k modelování a simulaci přírodních, technických a společenských procesů a k jejich implementaci v různých oborech. Zde je vidět možnost propojení statistických znalostí a práce se softwarem ke zpracování dat, které bylo nastíněno, i přesahů v RVP ZŠ.

4.3. Přesahy statistické gramotnosti do průřezových témat na základních školách a gymnáziích

Přesahy statistické gramotnosti se objevují nejen ve vzdělávacích oblastech, ale i v průřezových tématech. Ta reprezentují v RVP okruhy aktuálních problémů současného světa a jsou běžně zařazována do výuky. Právě v průřezových tématech, ačkoliv to není na první pohled patrné, se můžeme setkat s aplikací statistických metod. Statistická gramotnost se prolíná nejen se vzdělávacími oblastmi, ale i s průřezovými tématy. Ta jsou na základní škole a gymnáziu stejná vyjma tématu *Výchova demokratického občana*, které je vyučováno pouze na základní škole. Z průřezových témat nabízí k rozvoji statistické gramotnosti prostor především tato témata:

Průřezové téma **Výchova demokratického občana** v rámci tematického okruhu *Formy participace občanů v politickém životě* objasňuje volební systémy a politiku na různých úrovních. V rámci témat *parlamentní, krajské a komunální volby*, obec jako základní jednotka samosprávy státu je vhodné

zařadit v rámci výuky statistiky samostatnou nebo naopak skupinovou práci, kdy žáci mají za úkol zjistit informace o výsledcích voleb a na tyto výsledky poté aplikovat získané vědomosti a dovednosti.

Výchova k myšlení v evropských a globálních souvislostech rozvíjí a integruje základní vědomosti potřebné pro porozumění sociálním a kulturním odlišnostem mezi národy. Dále rozvíjí schopnost srovnávat projevy kultury v evropském a globálním kontextu a schopnost nacházet společné znaky a odlišnosti a hodnotit je v širších souvislostech. Jednou z možností aplikace statistiky je srovnání demografických údajů zemí Evropské unie s výsledky České republiky a to v oblastech ekonomických, kulturních i společenských.

Multikulturní výchova se zabývá poměrně širokým spektrem problémů, ať už kulturními rozdíly, etnickým členěním společnosti, lidskými vztahy a principy solidarity. Statistiku tak využijeme např. k mapování rozdílných sociálních jevů.

Environmentální výchova využívá metody popisné statistiky především v rámci přírodovědných předmětů (fyziky, chemie, přírodopisu, zeměpisu a výchovy ke zdraví). Prakticky ve všech těchto předmětech se žáci setkají s případy, ve kterých potřebují zpracovat naměřená, získaná nebo vypočtená data.

Mediální výchova se zabývá dnes velice aktuálním tématem médií a jeho vlivu na chování jedince a společnosti. Média dnes ovlivňují celkový životní styl a kvalitu života. Přitom sdělení, jež jsou médiu nabízena, jsou často zavádějící a vyznačují se nepřesnými, neseriózními a nepodloženými zdroji informací. Správné vyhodnocení takovýchto sdělení a jejich vztahu k realitě vyžaduje značnou průpravu [7]. Schopnost žáka orientovat se v datech uváděných v různých grafech, tabulkách, rozumět těmto sdělením, správně je interpretovat a kriticky hodnotit velice úzce souvisí se statistickou gramotností. V souladu s požadavky současné společnosti je třeba zdůraznit kritický postoj k interpretacím statistických sdělení, poukázat na záměrné nekorektní zneužívání výsledků statistických průzkumů a manipulaci v reklamách.

Osobnostní a sociální výchova obecně formuje studijní dovednosti, které jsou nezbytné pro další vzdělávání. Přesah statistiky do tohoto průřezového tématu není jednoznačný.

5. Získané kompetence na základních školách a gymnáziích

Klíčové kompetence představují soubor vědomostí, dovedností, schopností, postojů a hodnot, které jsou důležité pro osobní rozvoj jedince, jeho aktivní zapojení do společnosti a budoucí uplatnění v životě [1]. I vzdělávání

ve vzdělávací oblasti Matematika a její aplikace směřuje k utváření a rozvíjení klíčových kompetencí. V etapě základního a středního vzdělávání jsou za klíčové považovány: kompetence k učení; kompetence k řešení problémů; kompetence komunikativní; kompetence sociální a personální; kompetence občanské; kompetence pracovní [1, str. 14]. Tabulka 2 zachycuje kompetence získávané na základních školách a gymnáziích. Tabulky kompetencí v RVP ZŠ a RVP G obsahují všechny schopnosti a dovednosti, kterých by žáci na dané úrovni měli dosáhnout, v Tabulce 2 jsou však vybrány pouze ty, které souvisí s rozvojem statistické gramotnosti. Pro její rozvoj jsou nezbytné zejména čtyři klíčové kompetence, jež by si měli žáci osvojit v průběhu studia na základních školách a gymnáziích, a to kompetence k řešení problémů, kompetence komunikativní a kompetence občanská a pracovní.

Tabulka 2: Shrnutí osvojených pojmů a dovedností ve vzdělávací oblasti Matematika a její aplikace na základních školách a gymnáziích

Kompetence	Základní školy	Střední školy
K učení	<ul style="list-style-type: none"> • Vyhledává a třídí informace a na základě jejich pochopení, propojení a systematizace je efektivně využívá v procesu učení, tvůrčích činnostech a praktickém životě, • vytváří si komplexnější pohled na matematické, přírodní, společenské a kulturní jevy, • samostatně pozoruje a experimentuje, získané výsledky porovnává, kriticky posuzuje a vyvozuje z nich závěry pro využití v budoucnosti. 	<ul style="list-style-type: none"> • Kriticky přistupuje ke zdrojům informací, informace tvořivě zpracovává a využívá při svém studiu a praxi, • kriticky hodnotí pokrok při dosahování cílů svého učení a práce, přijímá ocenění, radu i kritiku ze strany druhých, z vlastních úspěchů i chyb čerpá poučení pro další práci.

<p>Komunika- tivní</p>	<ul style="list-style-type: none"> • Rozumí různým typům textů a záznamů, obrazových materiálů a jiných informačních a komunikačních prostředků, přemýšlí o nich, reaguje na ně a tvořivě je využívá ke svému rozvoji a k aktivnímu zapojení se do společenského dění. 	<ul style="list-style-type: none"> • Využívá dostupné prostředky komunikace včetně symbolických a grafických vyjádření informací různého typu, • používá s porozuměním odborný jazyk a symbolická a grafická vyjádření informací různého typu, • rozumí sdělením různého typu v různých komunikačních situacích, správně interpretuje přijímaná sdělení a věcně argumentuje.
<p>K řešení problémů</p>	<ul style="list-style-type: none"> • Vyhledá informace vhodné k řešení problému, nachází jejich shodné, podobné a odlišné znaky, využívá získané vědomosti a dovednosti k objevování různých variant řešení, • užívá při řešení problémů logické, matematické a empirické postupy, • ověřuje prakticky správnost řešení problémů a osvědčené postupy aplikuje při řešení obdobných nebo nových problémových situací, • kriticky myslí, činí uvážlivá rozhodnutí, je schopen je obhájit, uvědomuje si zodpovědnost za svá rozhodnutí a výsledky svých činů zhodnotí. 	<ul style="list-style-type: none"> • Vytváří hypotézy, navrhuje postupné kroky, zvažuje využití různých postupů při řešení problému nebo ověřování hypotézy, • uplatňuje při řešení problémů vhodné metody a dříve získané vědomosti a dovednosti, využívá analytického a kritického myšlení, • kriticky interpretuje získané poznatky a zjištění a ověřuje je, pro své tvrzení nachází argumenty a důkazy, formuluje a obhajuje podložené závěry, • nahlíží problém z různých stran, • zvažuje možné klady a zápory jednotlivých variant řešení, včetně posouzení jejich rizik a důsledků.

Občanské	<ul style="list-style-type: none"> • Rozhoduje se zodpovědně podle dané situace, poskytne dle svých možností účinnou pomoc a chová se zodpovědně v krizových situacích i v situacích ohrožujících život a zdraví člověka, • chápe základní ekologické souvislosti a environmentální problémy, respektuje požadavky na kvalitní životní prostředí, rozhoduje se v zájmu podpory a ochrany zdraví a trvale udržitelného rozvoje společnosti. 	<ul style="list-style-type: none"> • Chová se informovaně a zodpovědně, • posuzuje události a vývoj veřejného života, sleduje, co se děje v jeho bydlišti a okolí, zaujímá a obhajuje informovaná stanoviska a jedná k obecnému prospěchu podle nejlepšího svědomí.
Sociální a personální	—	<ul style="list-style-type: none"> • Rozhoduje se na základě vlastního úsudku, odolává společenským i mediálním tlakům.
Pracovní	<ul style="list-style-type: none"> • Činí podložená rozhodnutí o dalším vzdělávání a profesním zaměření, • orientuje se v základních aktivitách potřebných k uskutečnění podnikatelského záměru a k jeho realizaci, chápe podstatu, cíl a riziko podnikání, rozvíjí své podnikatelské myšlení. 	<ul style="list-style-type: none"> • Získává a kriticky vyhodnocuje informace o vzdělávacích a pracovních příležitostech, využívá dostupné zdroje a informace při plánování a realizaci aktivit, • posuzuje a kriticky hodnotí rizika související s rozhodováním v reálných životních situacích, • vyhledává a kriticky posuzuje příležitosti k uskutečnění podnikatelského záměru s ohledem na své předpoklady, realitu tržního prostředí a další faktory.

Kompetence k učení souvisí se statistickou gramotností pouze okrajově, avšak je třeba tuto souvislost popsat. Žáci by měli umět pracovat s informacemi a kriticky je hodnotit. Hlavním důvodem je zejména irelevantnost některých informací publikovaných na internetu a v médiích. Právě komplexní pohled na celou problematiku a kritické smýšlení o informacích může žákům značně ulehčit učení, neboť při vhodném výběru studijních materiálů mohou

zdrojům věřit a nemusí se k problému neustále vracet. Příkladem mohou být referáty, které jsou zpravidla inspirovány zdroji z internetu a jiných médií. Pokud žák důvěřuje informačně nekvalitním zdrojům, je přínos ze zpracování referátu zanedbatelný. **Kompetence k řešení problému** je z hlediska statistické gramotnosti zásadní, neboť žák kompetentní k řešení problémů rozumí sdělením různého typu v různých komunikačních situacích a správně interpretuje. Tato kompetence je zcela nezbytná pro statistickou gramotnost, neboť sdělení skrze čísla a statistiky jsou veřejností často dezinterpretována. Žák kompetentní k řešení problémů navíc využívá empirické postupy k ověřování svých hypotéz, které umí obhájit před publikem. Dále umí zvážit různé varianty řešení problému a posoudit jejich rizika.

Kompetence komunikativní souvisí zejména s porozuměním žáka různým vyjádřením informací například v grafické nebo tabulkové podobě. Tyto typy informací umí i samostatně vytvořit. Tyto schopnosti úzce souvisí s prezentací získaných informací. Význam této kompetence pro statistickou gramotnost tkví především v tom, že jsou statistické informace velice často publikovány formou grafů a tabulek. Mezi další kompetence patří **kompetence sociální a personální**, jež umožňuje žákovi se rozhodovat na základě vlastního úsudku a odolávat společenským a mediálním tlakům. V neposlední řadě souvisí statistická gramotnost i s **kompetencí občanskou**, neboť každý občan se na základě statistik rozhoduje. Například v předvolebním období probíhají výzkumy veřejného mínění a jsou publikovány nejrůznější statistiky, na základě nichž občan usuzuje. Předvolební období je na statistiky velice bohaté – mediální agentury se ohání odhady volebních preferencí, politici výši schodku státního dluhu anebo výši průměrného platu. Právě porozumění těmto sdělením je nezbytnou součástí statistické gramotnosti, neboť při pochopení vzniku takto publikovaných čísel občan může sám kriticky posoudit jejich relevantnost. Právě v průběhu studia na gymnáziích žáci nabývají plnoletosti a poprvé smí využít své volební právo.

Žák s osvojenou pracovní kompetencí získá a kriticky zhodnotí informace při plánování a realizaci například podnikatelských aktivit. Podnikatelé často musí nahlížet na trh skrze nejrůznější statistiky publikované českým statistickým úřadem. Tato čísla jim dávají informace o možných podnikatelských příležitostech nebo o ekonomickém vývoji jejich podnikatelského odvětví.

6. Závěr

Cílem této práce bylo zjistit, do jaké míry se objevuje statistika v RVP ZŠ a RVP G. K jeho naplnění bylo využito obsahové analýzy těchto dokumentů. V rámci této analýzy bylo zjištěno několik zásadních poznatků:

- 1) Statistická gramotnost je upevňována zejména v oblasti *Matematika a její aplikace*. Praktické využití statistiky není nikde v RVP explicitně uvedeno a tudíž se lze domnívat, že se do výuky dostane jen málokdy. Během studia na základní škole a gymnáziu si žáci osvojí pojmy shrnuté v Tabulce 1. V průběhu základní školy si žáci osvojí pojmy kvantitativní a kvalitativní znak, průměr, modus a medián. Mimo toho se naučí číst v grafech a porozumět jejich tvorbě. Na gymnáziích se navíc vyučují i pojmy náhodný jev, pravděpodobnost, statistický soubor, percentil, kvartil, směrodatná a mezikvartilová odchylka. Tyto pojmy už plně zahrnují charakteristiky souboru – jak míry polohy, tak i míry variability. Úplně zde však chybí základy pravděpodobnosti, kde by si žáci mohli na praktických příkladech uvědomit, co je náhodnost a variabilita (např. házení kostkami, atd.). Žáci by si na gymnáziu měli osvojit dovednosti srovnávání souborů na základě odlišnosti charakteristik souborů. Mimo statistického chápání by však měli pochopit i přesahy do reálných situací. Například v rámci průřezového tématu *Mediální výchova* by si žáci měli uvědomit, že média často využívají statistiky (průměry, mediány) jako argumenty. Žáci pak mají kriticky zhodnotit využívání těchto statistik a vzhledem k tomu, že jim rozumí a znají je, posoudit i relevantnost jejich vzniku.
- 2) Během studia na základní škole a gymnáziu si žáci osvojí mnoho dovedností. Pro přehlednost jsou shrnuty v Tabulce 2. Mezi zásadní patří kompetence k řešení problémů. Žák vybaven touto kompetencí vytváří hypotézy, navrhuje postupné kroky a zvažuje využití různých postupů při řešení problému, dále uplatňuje při řešení problémů vhodné metody a dříve získané vědomosti a dovednosti. Využívá také analytického a kritického myšlení, což je nezbytné pro další rozvoj statistické gramotnosti.
- 3) Statistická gramotnost se prolíná nejen skrze různé vzdělávací oblasti ale i skrze průřezová témata. Například v rámcových vzdělávacích programech je statistika schována zejména v oblasti *Matematika a její aplikace*, avšak prolíná se i s výukou ICT předmětů i přírodovědnými a dalšími předměty. Statistiku můžeme najít nejen ve vzdělávacích oblastech, ale i v průřezových tématech.

Na vysoké školy tedy nastupují studenti, absolventi základních škol a gymnázií, kteří by měli znát základní pojmy popisné statistiky, které si osvojili v průběhu studia na základní škole a gymnáziu. Mimo těchto pojmů si také průběžně osvojili různé schopnosti a rozvíjeli si klíčové kompetence, což znamená, že studenti prvních ročníků vysokých škol by měli umět popsat zá-

kladní soubory dat pomocí měr polohy a variability, kriticky zhodnotit a interpretovat získané poznatky a ověřit je. Pro svá tvrzení by měli najít nejen slovní, ale i číselné argumenty. V rámci tvorby měr polohy a variability by měli umět využívat i tabulkové kalkulátory. Právě tato úroveň získaných znalostí a kompetencí souvisejících se statistickou gramotností je zcela zásadní pro studium na vysoké škole.

Vysokoškolské studenty čeká psaní bakalářské a diplomové práce, a proto budou statistiku a statistické metody potřebovat. Pokud ovládají pojmy nabyté v průběhu studia na základních školách a gymnáziích a osvojili si klíčové kompetence, jsou plně připraveni ke studiu kurzu základů statistiky, který tyto kompetence bude dále rozvíjet a připravovat je nejen na psaní závěrečné práce, ale i na případnou vědeckou kariéru. V případě, že student v závěrečné práci nevyužije statistických metod, ani neplánuje vědeckou kariéru, je kurz základů statistických metod vhodný, neboť, jak již bylo řečeno dříve, obklopují nás čísla a média i politici je velice často využívají jako argumenty. Jednou z možností jak se lépe orientovat v těchto argumentech je právě pochopení vzniku publikovaných statistik a tudíž i možnost je kriticky zhodnotit.

Otázkou k diskuzi zůstává, zda studenti nastupující na vysoké školy mají znalosti deklarované rámcovými vzdělávacími programy a zda tedy v tomto ohledu plní základní školy a gymnázia svou úlohu.

Poděkování

Výzkum byl podpořen Grantovou agenturou Univerzity Karlovy v Praze (projekt č. 341115).

Literatura

- [1] Balada, J.: *Rámcový vzdělávací program pro gymnázia: RVP G*. Výzkumný ústav pedagogický v Praze, Praha, 2007.
- [2] Ben-Zvi, D., Garfield, J. B. (eds.): *The challenge of developing statistical literacy, reasoning, and thinking*. Kluwer academic publishers, Dordrecht, 2004.
- [3] Garfield, J.: Thinking about statistical reasoning, thinking and literacy. *First Annual Roundtable on Statistical Thinking, Reasoning and Literacy (STRL-1)*, 1999.
- [4] Hybšová, A.: Statistická gramotnost studentů učitelství biologie na pedagogických fakultách českých univerzit. In: *Profesní příprava učitelů přírodovědných oborů: sborník materiálů z vědecké konference s mezinárodní účastí: Olomouc, 3. 4. 2013*. UPOL, Olomouc, 2013.

- [5] Hybšová, A.: Statistická gramotnost a její význam v pedagogice a oborových didaktikách. In: *Aktuální problémy pedagogiky ve výzkumech studentů doktorských studijních programů*. UPOL, Olomouc, 2014.
- [6] Hybšová, A.: Statistical literacy in teaching of natural science. In: *Turning data into knowledge: New opportunities for statistics education*. V tisku.
- [7] Pernicová, T.: Využití programu MS Excel ve výuce matematiky na ZŠ: diplomová práce. Brno: Masarykova univerzita, Fakulta pedagogická, Katedra matematiky, 2013.
- [8] *Rámcový vzdělávací program pro základní vzdělávání*. Upravené vydání se změnami účinnými od 1. 9. 2013. Národní ústav pro vzdělávání, Praha, 2013. URL: <http://www.nuv.cz/cinnosti/kurikulum-vseobecne-a-odborne-vzdelavani-a-evaluace/ramcove-vzdelavaci-programy/upraveny-rvp-zv>
- [9] Saxl, I.: Statistické myšlení a jeho výuka. In: *Pravděpodobnost a statistika na střední škole. Sborník prací didaktického semináře pořádaného MFF UK v Praze*. Matfyzpress, Praha, 2005, pp. 1–16.
- [10] Stehlíková, B., Markechová, D.: Štatistika – vstupná brána poznania. In: *Forum Statisticum Slovacum 2/2011*, 171 stran.
- [11] Wallman, K. K.: Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association* 88 (421), pp. 1–8.

ZEMŘEL ZDENĚK ROTH

ZDENĚK ROTH PASSED AWAY

Marek Malý, Bohumír Procházka

E-mail: marek.maly@szu.cz, bohumir.prochazka@szu.cz

Nedlouho po svých 85. narozeninách náhle zemřel renomovaný statistik Ing. Zdeněk Roth, CSc., dlouholetý pracovník Státního zdravotního ústavu a bývalý předseda České statistické společnosti. Ztrácíme v něm váženého člověka, vynikajícího statistika, nepřehlédnutelnou osobnost statistické komunity. Po 60 let aktivně pracoval v oblasti analýzy biologických a lékařských dat a statistice se věnoval do posledních chvil. Chtěli bychom zde připomenout některé okamžiky z jeho bohatého života.

Po maturitě na Benešově reformním reálném gymnasiu v Praze XIX. a ročním odkladu kvůli onemocnění tuberkulózou začal studovat statistiku na fakultě speciálních nauk ČVUT, po reformě vysokých škol studium zakončil na Matematicko-fyzikální fakultě UK, kde v roce 1954 získal diplom inženýra ekonomie se specializací matematická statistika.

Pracoval nejprve jako statistik ve farmakologickém oddělení Výzkumného ústavu pro farmacii a biochemii v Praze (1954–1960), kde se zabýval hlavně statistickým vyhodnocováním dat z experimentů a klinických pokusů. Poté vedl statistické oddělení v Ústavu hygieny práce a chorob z povolání (1961–1971), kde témata statistických analýz sahala od ergonomie a psychologie přes toxikologii k epidemiologii.

V době, kdy mu bylo teprve 30 let, připravil jako hlavní autor společně s Marcelem Josífkem, Vladimírem Malým a Václavem Trčkou knihu *Statistické metody v experimentální medicíně*, která se zabývá především problematikou vyhodnocování farmakologických a biologických pokusů. Dodnes cenná kniha vyšla v Státním zdravotnickém nakladatelství v roce 1962, její uvažovaná reedice už nebyla umožněna.

Výsledkem zájmu Ing. Rotha o problémy obecné farmakologie a toxikologie byla rovněž jeho kandidátská práce „Stochastické modely receptorové teorie účinku drog a testy, jejich kompetivní interakce a receptory“, kterou obhájil v roce 1968 na Katedře matematické statistiky a pravděpodobnosti MFF UK (školitel prof. L. Truksa, vedlejší školitel MgMat. M. Josífk) a získal titul kandidáta fyzikálně-matematických věd. Své zkušenosti z této oblasti využil i jako člen Lékopisné komise Ministerstva zdravotnictví, podílel se na přípravě třetího vydání Československého lékopisu (1970).

V 60. letech 20. století využil Ing. Roth výjimečné příležitosti pracovat po dva roky (1965, 1969) v oddělení Sociálního lékařství a klinické epide-

miologie nemocnice sv. Tomáše v Londýně u prof. W. W. Hollanda nejprve jako stážista a později jako asistent odpovědný za plánování a analýzu epidemiologických studií, včetně využití tehdejších samočinných počítačů. V roce 1967 byl zvolen členem Mezinárodní epidemiologické společnosti a aktivně se podílel na její činnosti, na jejím pátém kongresu v srpnu 1968 měl jednu z hlavních přednášek. Díky této činnosti se stal členem komise expertů zdravotnické statistiky WHO (1971–1980).

Na počátku 70. let po ustavení Institutu hygieny a epidemiologie, dnešního Státního zdravotního ústavu, došlo ke sjednocení statistiků, grafiků, programátorů a počítačových techniků z několika dílčích pracovišť v Oddělení matematické statistiky a programování, které Ing. Roth výrazně formoval a v letech 1971–1995 vedl. Zvládl zde se spolupracovníky ohromný objem práce při analýze dat a podpoře výzkumu z rozličných oblastí biologie a medicíny. Ve Státním zdravotním ústavu pracoval Ing. Roth, především na různých grantových úkolech, i po odchodu do důchodu, až do roku 2015.

Dlouhodobé vedlejší úvazky pojily Zdeňka Rotha především s Výzkumným ústavem psychiatrickým, pozdějším Psychiatrickým centrem Praha (od 1962), dále s oddělením toxikologie Vojenského výzkumného a doškolovacího ústavu v Hradci Králové (1960–1985), Institutem pro další vzdělávání lékařů a farmaceutů (1973–1993) a 3. lékařskou fakultou UK (1990–1996).

Aktivní byl i v práci pro statistickou komunitu. Dlouhodobě zastával pozici předsedy sekce zdravotnické statistiky při Společnosti sociálního lékařství Lékařské společnosti J. E. Purkyně. V letech 1995–2001 byl po tři volební období historicky třetím předsedou České statistické společnosti a na práci jejího výboru se podílel až do roku 2013. Jeho snahou vždy bylo pořádání seminářů umožňujících výměnu nejnovějších informací a odborný kontakt mezi členy společnosti.

Zdeněk Roth měl vždy cit pro počítačové zpracování dat a od let, kdy musel pečlivě a s hlubokým porozuměním problematiky rozepisovat plány statistických analýz pro výpočtářky (chyba tehdy znamenala třeba několik dní práce), rychle přešel k prvním sálovým počítačům, jako byl Minsk 22 či později Hewlett-Packard, na nichž připravoval nejen programy nestandardních matematicko-statistických analýz, ale i značně odlišných úloh, jako bylo zpracování mezd. Výborně dokázal realizovat svoje statistické znalosti v počítačových programech, které sám vytvářel. Rychle se adaptoval na nový typ techniky, a tak od autorských programů došel až k současným počítačům a komerčním statistickým programům. Standardní naprogramované algoritmy často rozvíjel a doplňoval netradičními a originálními postupy.

Ing. Roth je autorem či spoluautorem stovek odborných publikací vzniklých v průběhu 60 let jeho statistické praxe, první vyšla v roce 1956 a poslední,

na níž se podílel se svou dcerou, v roce 2015. Souhrnný seznam prací není momentálně k dispozici, nicméně již k roku 1970 bylo zaznamenáno 60 jeho časopiseckých publikací. Přitom vždy říkal, že ho podstatně víc baví řešení problémů vymýšlet, než jej poté formulovat do článku.

Konzultace, vymýšlení řešení a statistické výpočty pro biology a lékaře byly hlavní doménou práce Ing. Rotha. Biologové a lékaři si jej velmi vážili právě proto, že byl schopen se s nimi domluvit a na základě mnohdy nepřiliš jasného zadání navrhnout adekvátní model, často poté, co i precizně zformuloval zkoumané hypotézy. Jeho řešení se vyznačovala jak kvalifikovaným statistickým přístupem, tak pochopením pro realitu citem pro interpretací získaných výsledků. Zároveň byl vždy ochoten se o své rozsáhlé znalosti podělit s kolegy statistiky a přispět radou k řešení jejich problémů. Často a rád přednášel na různých fórech, a to jak o výsledcích provedených analýz, tak o statistice jako takové. Otázky výuky biostatistiky považoval za velmi důležité a sám se dlouhodobě podílel na výuce studentů a na prohlubování znalostí lékařů v oblasti užití statistických metod v medicíně a epidemiologii.

Zdeněk Roth nám byl vzorem nejen v tom, jak řešit odborné problémy, ale i tím, jak žít plnohodnotný život navzdory zdravotním omezením, jak se dívat dopředu a neohlížet zpět. Vždy nás fascinovala jeho vitalita, houževnatost a nezdolný optimismus. Díky své šťastné povaze si nepřipouštěl mnohé problémy a jiné snadněji překonával. Charakteristické například bylo, když se vydal na konferenci Robust v Lednici hromadnou dopravou přesto, že měl zlomenou nohu sešroubovanou zvenku dráty. Večer jsme jej našli spokojeně sedět v recepci hotelu, kde se konference konala.

Skvělým způsobem se mu dařilo tmelit pracovní kolektiv i mimo pracovní prostředí. Měl velký rozhled nejen ve statistice, ale i v mnoha dalších oblastech. Byl pro nás skutečným přítelem, se kterým jsme sdíleli jeho nadšení pro práci na zahradě, chov včel a pro hudbu. Byla radost s ním diskutovat o hudebních zážitcích – sám byl výborný klavírista a neúnavný sběratel hudebních nahrávek, zejména jazzu, swingu a vážné hudby.

Marek Malý, Bohumír Procházka

JAK JSEM SE SETKÁVAL S ING. ZDENĚKEM ROTHEM

MEETINGS WITH A STATISTICIAN MR. ZDENĚK ROTH

Josef Tvrdík

E-mail: josef.tvrdik@osu.cz

K prvnímu setkání došlo někdy kolem roku 1980 a bylo to setkání čtenáře se statistickou monografií, jejímž prvním autorem byl Zdeněk Roth. Pracoval jsem tehdy v Krajské hygienické stanici (KHS) v Ostravě a jedna paní doktorka, která se zabývala pracovním lékařstvím, mi přinesla svá data a knihu autorů Roth, Josífko, Malý, Trčka, s tím, že mám spočítat probitovou regresi. A ještě dodala něco v tom smyslu, ať se nevymlouvám, že to nejde, protože ona to počítala na cvičení statistiky u prof. Komendy na LF v Olomouci. Tak mi nezbylo, než to nastudovat. Byl jsem příjemně překvapen, jak srozumitelně byla tato partie v knize vysvětlena. Úlohu jsem byl schopen docela rychle vyřešit ke spokojenosti paní doktorky. Říkal jsem si, že autoři této knihy musí být velmi moudří a patrně i staří muži, vždyť kniha vyšla v době, kdy já začínal studovat VŠCHT a o statistice neměl ani mlhavou představu. Mnohem později jsem pány Rotha a Josífka potkal a zjistil, že to jsou opravdu muži moudří, ale zdaleka ne tak staří, jak jsem to podle roku vydání knihy odhadoval.

Druhé setkání někdy v polovině 80. let bylo telefonické, kdy jsem Ing. Rotha slyšel, ale neviděl. V KHS se zpracovával informační systém o přenosných onemocněních a souhrnná data se pak posílala na magnetické pásce do IHE, kde se pak vydávaly nějaké celostátní přehledy. Kolegové, kteří se o tento systém starali, byli na dovolené nebo na delší služební cestě, a tak mě vybavili pokyny, jaké programy mám spustit, nahrát data, magnetickou pásku poslat do IHE a ještě jim zatelefonovat, že páska je na cestě. Provedl jsem, zavolaal na zadané číslo a v telefonu se ozvalo „Roth“. Lehce jsem se zděsil, že s tímto starcem se bude obtížné domluvit o takovém výdobytku techniky, jako je magnetická páska. K mému příjemnému překvapení vzápětí se Ing. Roth zasvěceně dotázal, zda je páska nahrána s labely nebo bez labelů. Když jsem oznámil, že bez labelů, pochválil nás, že usnadňujeme práci s přečtením na jiném typu počítače, poděkoval a popřál hodně zdaru. Až daleko později jsem se dozvěděl, že Zdeněk byl s počítači jedna ruka a že předtím i po tom naprogramoval ve Fortranu spoustu statistických modulů, ze kterých si sestavoval programy pro statistickou analýzu právě zpracovávaných dat.

Pak už jsme potkávali opravdu živě na Robustech, valných hromadách a výborech Statistické společnosti, Statistických dnech, Stakanech a podobných akcích. Při těchto příležitostech jsme prodebatovali spoustu věcí statistických i nestatistických, vypili nejednu sklenku vína a přezpívali mnoho písniček (Zdeněk velice často u piana). V mnoha důležitých věcech jsme se naprosto shodovali, např. na prospěšnosti koupání v přírodě bez plavek, neboť to prospívá tělu i duchu. Jen jsem byl měkčí v tom, že na podzim jsem končil o dost dříve a na jaře začínal o trochu později než Zdeněk. S potěšením vzpomínám na první večer na Stakanu v Cikháji, kdy se nás několik po úmorné cestě v horkém dnu zajelo vykoupat do proudů pod výpustí v té době prázdného Velkého Dářka (tam jsme ale bez plavek nemohli, to by bylo pohoršení početné veřejnosti, za které by nás tehdy Marie Budíková mohla ze Stakanu vyloučit). Po večeři jsme zasedli k vínu pod hvězdnou letní oblohu v dobré společnosti, v níž hráli prim prof. Komenda a Zdeněk Roth. Bylo radost poslouchat debatu a zpívat s nimi lidovky. Myslím, že Zdeňkovi ani nevadilo, že venku nemá piano.

Jedno z posledních našich setkání „na živo“ bylo na Robustu 2008 v Pribylině. Ve středečním odpoledni obvyklého turistického výletu jsem byl coby chromý horské turistiky neschopen, Zdeňkovi už také nohy úplně dobře nesloužily, a tak nám slovenští kolegové poradili, ať si zajedeme do aquaparku na Oravici. Ještě se přidaly Hanka Řezanková a Marta Žambochová, ale ty pak daly přednost turistické vycházce do krásného okolí Oravice, a tak nemohou dosvědčit, jak se dědci v aquaparku vyblbnuli. Kromě plavání tam byl i docela vysoký tobogán. Zdeněk předtím na tobogánu nikdy nejel, zatímco já jsem měl už za sebou několikaletou praxi s vnuky. Bylo jasné, že takovou příležitost Zdeněk nezhodí. Nezbylo mi, než odložit dole berle a po zábradlí se rukama vyškrabat nahoru. Kluk Zdeněk byl jízdou na tobogánu nadšen a já musel po zábradlí šplhat nahoru opakovaně, protože jsem nemohl dopustit, abych jako mladší zaostal.

Měl jsem to štěstí setkávat se Zdeňkem Rothem, skvělým statistikem a velkým člověkem. Vděčně vzpomínám.

Josef Tvrdík

Obsah

Vědecké, odborné a přehledové články

Martin Branda, Jan Novotný, Asmund Olstad, Pavel Popela

Krátké pojednání o problému rozvrhování s předepsanými
časy prací a náhodnými prvky 1

Josef Ditrich

Použití logistické regrese pro malé datové výběry 8

Aneta Hybšová

Statistická gramotnost v rámcových vzdělávacích programech 18

Zprávy a informace

Marek Malý, Bohumír Procházka

Zemřel Zdeněk Roth 36

Josef Tvrdík

Jak jsem se setkával s Ing. Zdeňkem Rothem 39

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214. Časopis je na Seznamu recenzovaných neimpaktovaných periodik vydávaných v ČR, více viz server <http://www.vyzkum.cz/>.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in bulletin are published in English, Czech and Slovak languages.

Předsedkyně společnosti: prof. Ing. Hana ŘEZANKOVÁ, CSc., KSTP FIS VŠE v Praze, nám. W. Churchilla 4, 130 67 Praha 3, e-mail: hana.rezankova@vse.cz.

Redakce: prof. RNDr. Gejza DOHNAL, CSc. (šéfredaktor), prof. RNDr. Jaromír ANTOCH, CSc., prof. Ing. Václav ČERMÁK, DrSc., doc. Ing. Jozef CHAJDIK, CSc., doc. RNDr. Zdeněk KARPÍŠEK, CSc., RNDr. Marek MALÝ, CSc., doc. RNDr. Jiří MICHÁLEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. Ing. Iveta STANKOVIČOVÁ, Ph.D., doc. Ing. Josef TVRDÍK, CSc., Mgr. Ondřej VENCÁLEK, Ph.D.

Redaktor časopisu: Mgr. Ondřej VENCÁLEK, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, <http://www.statspol.cz/>.

DOI: 10.5300/IB, <http://dx.doi.org/10.5300/IB>
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vtištěno s laskavou podporou Českého statistického úřadu.