

# Posudek oponenta diplomové práce

Studijní program: **Kvantitativní metody v ekonomice**

Studijní obor: **Quantitative Economic Analysis**

Akademický rok: **2019/2020**

Název práce: **Logistic regression in R**

Řešitel: **Alicem Karaca**

Vedoucí práce: **Mgr. Milan Bašta, Ph.D.**

Oponent: **doc. RNDr. Ivana Malá, CSc.**

	Hlediska	Stupeň hodnocení
1.	Jasnost a srozumitelnost formulace tématu a cíle práce	<b>3</b>
2.	Míra naplnění cíle	<b>3</b>
3.	Logická stavba práce a vzájemná konzistence jednotlivých částí	<b>3</b>
4.	Rozsah a relevance popisu současného stavu	<b>3</b>
5.	Adekvátnost metod k řešení problému, korektnost výběru a aplikace	<b>2</b>
6.	Rozsah, hloubka a preciznost popisu výsledku	<b>3</b>
7.	Relevance a správnost diskuze výsledku	<b>3</b>
8.	Korektnost a relevance informačních zdrojů	<b>3</b>
9.	Gramatika, jazykový styl, terminologie a celková úprava práce	<b>2</b>

## Konkrétní připomínky a dotazy k práci:

The topic of the thesis is interesting, beneficial and suitable for the student of the quantitative master programme.

Unfortunately, the title of the work does not match the content, and it would be better to choose a title reflecting the practical problem being solved. The logistic regression model is used correctly, but it is only an instrument to solve a specific problem. The author deals with the topic carefully and is genuinely interested in the issue.

I appreciate that the author made a questionnaire survey, prepared the data and processed it in the program R.

The using of the program is at a reasonable level; the author has shown that he can formulate the problem, get the necessary data, construct a suitable statistical model and interpret the results.

It should be noted that the carefully described selection method does not provide a representative sample of any population. Therefore, statistical induction based on data is questionable, although the descriptive approach is correct. I wonder whether the sample size is sufficient for the inclusion of so many explanatory variables into the model.

There are too many strange, incomprehensible or at least inaccurate sentences as for example p.21: items 1.-3. , p. 22: "As it is known, the normal distribution of the regression analysis when the dependent variable is a categorical variable and the assumptions that the error terms are constant variance are ignored." or "Logistic regression model Equation 1.2. is indicated by. When  $X = x$  according to this relation,  $Y$  is the probability of being  $1 - \pi$ ." in the text.

The statistical notation is of poor quality.

In the discussion, the author should answer the following questions:

1. Could you describe in detail the construction of a confusion matrix? Could you define sensitivity and specificity from the Fig. 9.
2. Do you believe that your results could be generalized even though the sample is not representative?
3. What is wrong with the formula (2)?

**Závěr: Diplomovou práci doporučuji k obhajobě.**

Navrhovaná výsledná klasifikace práce: **3**

Datum: 15. 1. 2020

**doc. RNDr. Ivana Malá, CSc.**  
oponent práce