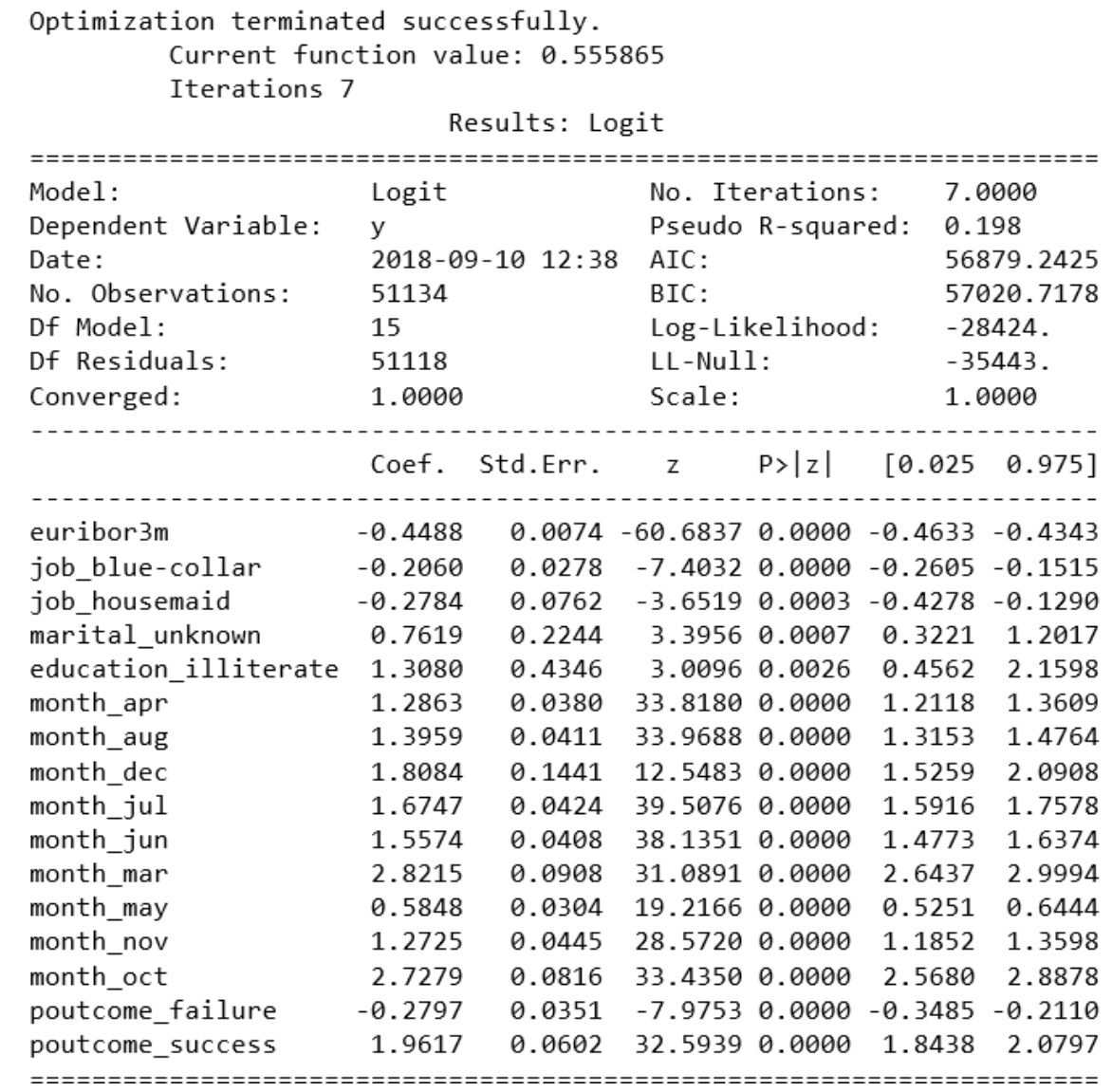
Lab 10

**Questions (only one correct answer)**

**Please take a look at the following table:**

**Table 1. Logistic regression results to predict whether a client subscribed a term deposit or not** (Li 2017)

****

**Q1. Which is the confidence interval presented in the table?**

1. **-0.4488 to 1.9617**
2. **0.025 to 0.975**

**Q2. What is the equivalent of confidence interval in a Bayesian logic (a confidence interval which takes into account the prior distribution, see** Statistics How to n.d.)**?**

1. **Trust interval**
2. **Credible interval**

**Q3. The highest probability for a client to subscribe is in:**

1. **April**
2. **March**
3. **October**

**Q4. Which of the following sentences is true?**

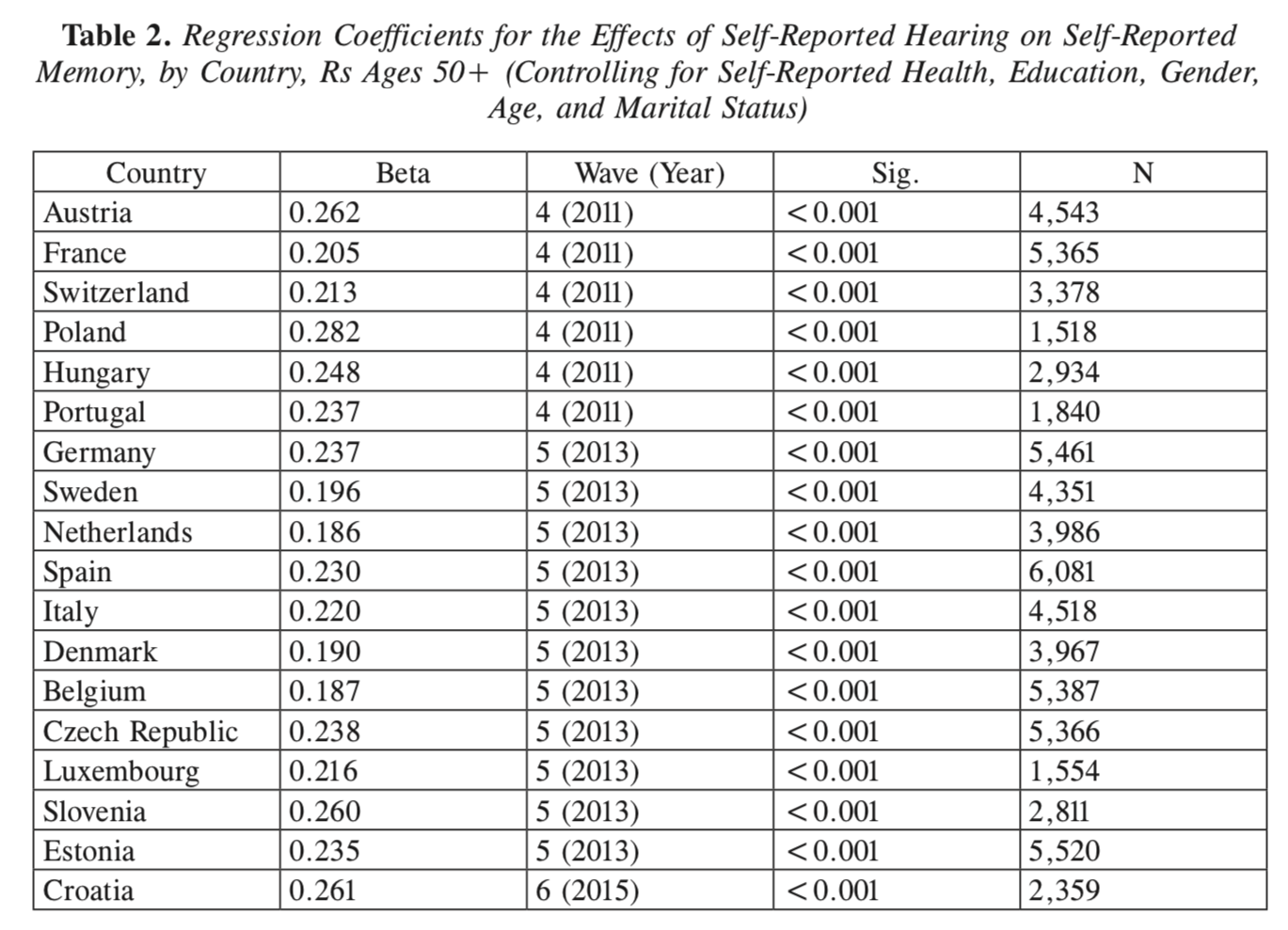
1. **The higher the Euribor rate is, the higher the chances for a client to subscribe.**
2. **The chances for a housemaid to subscribe are lower than those for other types of occupation.**
3. **Illiterate subjects are less likely to subscribe than subjects with higher levels of education.**

**Q5. Considering the results in Table 1, please evaluate the following sentence:**

***Approximately 20% of the variance in the dependent variable is explained by the model.***

1. **True**
2. **False**

**Take a look at the following table** (Ilinca and Cutler 2018)**:**

****

**Q6. Considering the results in Table 2, in which country is it more likely for hearing to have an influence on memory?**

1. **Spain**
2. **Poland**
3. **Austria**

**Q7. Considering the results in Table 2, please evaluate the following sentence:**

***The likelihood for hearing to influence memory is higher for women than for men.***

1. **True**
2. **False**

**Q8. Considering the results in Table 2, which is the country with the highest number of cases?**

1. **Spain**
2. **Czech Republic**
3. **Estonia**

**Programming**

**P1. Please take a look at the logistic regression described step by step by Susan Li** (2017) **and use her Jupyter notebook** (Li 2016) **in order to answer the following questions:**

1. **Test the model with only the first predictor (euribor3m). Does the coefficient value change? What about the Pseudo R-squared (i.e. the percentage of explained variance in the dependent variable – whether a client subscribed a term deposit or not – binary variable)? What is the accuracy of the logistic regression classifier in this case with only one predictor?**
2. **How many levels of analysis does the author use (individual level, company level, campaign level, country level)? How many levels are implied by the simple logistic regression? What suggestions do you have for the author?**
3. **To what extent does the substantive logic of results change if the model provides distributions of probability for each predictor instead of regression coefficients?**

Cited works:

Ilinca, Corina and Stephen J. Cutler. 2018. “Self‐Rated Hearing and Self‐Rated Memory II: A Cross‐Sectional Analysis of SHARE Data from 18 European Nations.” *Social Work Review* 17(4):127–35.

Li, Susan. 2017. “Building A Logistic Regression in Python, Step by Step.” *Towards Data Science*. Retrieved (https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8).

Li, Susan. 2016. “Jupyter Notebook: Machine-Learning-with-Python/Logistic Regression Balanced.Ipynb.” *Github*. Retrieved (https://github.com/susanli2016/Machine-Learning-with-Python/blob/master/Logistic Regression balanced.ipynb).

Statistics How to. n.d. “Credible Interval: Simple Definition.” Retrieved (https://www.statisticshowto.datasciencecentral.com/credible-interval/).