1. **The final paper should consist of**

**For example, examining a data set related to education or health using statistical models etc... Impact of some charecteristics on our model... etc. ( With R Language )..**

**(Time series models, Models with Binary Dependent Variables, Ordered Choice Models, Models for Count Data, Multinomial Logit, Conditional Logit, Limited Dependent Variables (tobit), Panel Data Models .... etc)**

**Note: Just one model above for should be used for the real data set belonging to the problem definition you find. Whichever model is chosen, extra requirements must also be do. ( Page 2, 3 and 4 )**

**It is enough to apply one of these models. My prefer is "Models with Binary Dependent Variables" but you can change.**

**The information on the page must be written in the order specified in the blue part.**

**The problem and hypothesis should be clearly explained.**

**The method to be chosen for the problem should be briefly mentioned.**

**The data set must be obtained from an accessible website. (worldbank, kaggle, etc ...)**

**Whether we accept the hypothesis or not should be stated for reasons. ( conclusion part )**

**(You can already find these extra special requirements for each model below)**

**If we accept this model, the additional requirements requested from me during the use of this model are as under red heading. ( Page 2, 3 and 4 )**

1. **The final paper should consist of**

**a) Title, Authors, Date**

**b) Abstract**

**c) Introduction**

**d) Literature review**

**e) Data**

**f) Method/Model**

**g) Results**

**h) Findings**

**i) Bibliography**

**j) Appendix**

1. Formal language
2. Abstract should comprise the main aims of the paper, short description of the method used, and main findings. Cf: https://en.wikipedia.org/wiki/Abstract\_(summary) and https://www.springer.com/ (link)
3. Introduction. Describe the problem. Introduce the main and the secondary hypotheses. Explain the importance of the selected topic. https://www.springer.com/ (link 2)
4. Data. Describe your data. Where do they come from? Describe the data transformations. If any data was removed, give the reason. A plot, histogram, or some form of data visualisation might be attractive to a reader and might help to better understand your data.
5. Method/Model. General to specific approach.
6. Results. Verify your hypotheses. Please formulate an explanation why the hypothesis was re- jected, if any. Publication quality table is necessary.
7. Findings. Repeat the findings. The next possible ways of handling the topic/problem.
8. Appendix. R-code with comments.
9. Title, data, data description, one main hypothesis, one secondary hypothesis, literature should be sent until 12 April 2020 10:00.

METHODS SHOULD BE !

**Time series models – ARDL models (If this model is selected, the requirements are as follows.)**

a) stationarity tests;

b) general-to-specific approach to variables selection;

c) parameters interpretation; d) hypotheses verification;

e) Ramsey-Reset test - linear form of the relationship;  
f) Breusch-Pagan’s and White’s tests – homoscedasticity;

g) Breusch-Godfrey test – no autocorrelation;  
h) compare the final ARDL model with ARIMA model.

2. **Cointegration analysis** (If this model is selected, the requirements are as follows.)

a) stationarity tests;  
b) cointegration testing;

c) general-to-specific approach to variables selection;  
d) parameters interpretation - short-term and long-term part;

e) hypotheses verification;  
f) Ramsey-Reset test - linear form of the relationship;

g) Breusch-Pagan’s and White’s tests – homoscedasticity; h) Breusch-Godfrey test – no autocorrelation.

3. **Models with Binary Dependent Variables** ( If this model is selected, the requirements are as follows.)

1. a)  estimation of linear probability model (OLS with White’s robust matrix), logit model, and

probit model, selection of significant variables;

1. b)  choice between logit and probit on the basis of information criteria;
2. c)  general-to-specific method to variables selection;
3. d)  at least one nonlinear relationship (variable to a power) and interaction between variables;
4. e)  present the general model (LPM, logit, and probit), the final model (the specif model) in one quality table. If there is space, at least one intermediate model might be presented;
5. f)  calculation and interpretation of marginal effects for the final model (from the general-to- specific approach);
6. g)  calculation and interpretation of odds ratios;
7. h)  perform the linktest and interpret the result;
8. i)  interpretation of the appropriate R2 statistics (R2 McKelvey-Zavoina, count R2, and adju- sted count R2;
9. j)  hypotheses verification;
10. k)  perform the Hosmer-Lemeshow and alike tests.

1

4. **Ordered Choice Models** ( If this model is selected, the requirements are as follows.)

1. a)  estimate ordered probit and ordered logit, selection of the covariates;
2. b)  general-to-specific method to variables selection;
3. c)  at least one nonlinear relationship (variable to a power) and interaction between variables;
4. d)  present the general models (LPM, logit, and probit), the final model (the specif model) in one quality table. If there is space, at least one intermediate model might be presented;
5. e)  calculation and interpretation of marginal effects for the final model (from the general-to- specific approach);
6. f)  interpretation of the appropriate R2 statistics (R2 McKelvey-Zavoina, count R2, and adju- sted count R2;
7. g)  perform the linktest and interpret the result;
8. h)  perform the the Hosmer-Lemeshow test, the Lipsitz, and the Pulkstenis-Robinson tests;
9. i)  check the proportional odds assumption.

5. **Models for Count Data** ( If this model is selected, the requirements are as follows.)

1. a)  estimate Poisson model, negative binomial regression, Zero-Inflated Poisson Model;
2. b)  choose the most appropriate one from the three mentioned in point (a);
3. c)  at least one nonlinear relationship (variable to a power) and interaction between variables;
4. d)  general-to-specific method to variables selection;
5. e)  present the general models (Poisson model, negative binomial regression, Zero-Inflated Pois- son Model), the final model (the specific model) in one quality table. If there is space, at least one intermediate model might be presented; Poisson model, negative binomial regression, Zero-Inflated Poisson model;
6. f)  interpret the final model parameters;
7. g)  interpretation of the appropriate R2 statistics (R2 McKelvey-Zavoina, count R2, and adjusted count R2;
8. h)  perform the linktest and interpret the result;

6. **Multinomial Logit** ( If this model is selected, the requirements are as follows.)

1. a)  general-to-specific method for variables selection;
2. b)  at least one nonlinear relationship (variable to a power) and interaction between variables;
3. c)  calculation and interpretation of marginal effects for the final model (from the general-to- specific approach);
4. d)  present the general model, the final model (the specific model) in one quality table. If there is space, at least one intermediate model might be presented;
5. e)  perform the linktest and interpret the result.

2

7. **Conditional Logit** ( If this model is selected, the requirements are as follows.)

1. a)  general-to-specific method to variables selection;
2. b)  at least one nonlinear relationship (variable to a power) and interaction between variables;
3. c)  calculation and interpretation of marginal effects for the final model (from the general-to- specific approach);
4. d)  present the general model, the final model (the specific model) in one quality table. If there is space, at least one intermediate model might be presented;
5. e)  perform the linktest and interpret the result.

8. Limited Dependent Variables (tobit)

1. a)  general-to-specific method to variables selection;
2. b)  at least one nonlinear relationship (variable to a power) and interaction between variables;
3. c)  calculate and interpret three kinds of marginal effects;
4. d)  present the general model, the final model (the specific model) in one quality table. If there is space, at least one intermediate model might be presented;
5. e)  interpretation of the appropriate R2 statistics;
6. f)  check if the residuals are normally distributed;
7. g)  perform the linktest and interpret the result.

9. **Panel Data Models** ( If this model is selected, the requirements are as follows.)

1. a)  estimate model with fixed and random effects estimators, check whether individual effects are significant;
2. b)  perform the Hausman specification test;
3. c)  general-to-specific method to variables selection;
4. d)  at least one nonlinear relationship (variable to a power) and interaction between variables;
5. e)  diagnostic tests for the final model;
6. f)  interpret the final model parameters;
7. g)  present the general model, the final model (the specific model) in one quality table. If there is space, at least one intermediate model might be presented.