Data analytics, Review for Jhon Velasquez team.
Reviewed by: Szymon Lis, Filip Olszowski.
20.06.2022 r.

Final Results: **24/24**

Problem formulation [0-4 pts]:

- is the problem clearly stated
  1 pt – describing life expectancy of third world countries in 2015 year

- what is the point of creating model, are potential use cases defined
  1 pt – obtaining model which is useful in further studies for predicting life expectancy

- where do data comes from, what does it contain
  1 pt - https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

- is preprocessing step clearly described
  1 pt – yes, normalization of data using linear scaling method

Model [0-4 pts]

- are two different models specified
  1 pt - yes

- are difference between two models explained
  1 pt – yes, differences are explained

- is the difference in the models justified (e.g. does adding additional parameter makes sense?)
  1 pt - yes, aproaches and decision behind choosing model are clearly described

- are models sufficiently described (what are formulas, what are parameters, what data are required) [1 pt]
  1 pt – models are properly described

Priors [0-4 pts]

- Is it explained why particular priors for parameters were selected
  1 pt - yes, obtained from real data using online plotter

- Have prior predictive checks been done for parameters (are parameters simulated from priors make sense)
  1 pt – yes, prior predictive checks have been done and results are visible on histogram

- Have prior predictive checks been done for measurements (are measurements simulated from priors make sense)
  1 pt – yes, same as above

- How prior parameters were selected
  1 pt – calculated manually from data with sigma as 30% of the mean

Posterior analysis (model 1) [0-4 pts]

- were there any issues with the sampling? if there were what kind of ideas for mitigation were used
  1 pt - no issues with the sampling

- are the samples from posterior predictive distribution analyzed
  1 pt - yes, there is histogram as justification with description

- are the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)
  1 pt - data is quite consistent with posterior predictive samples, results visible on histogram

- have parameter marginal distributions been analyzed (histograms of individual parameters plus summaries, are they diffuse or concentrated, what can we say about values)
  1 pt – distribiutions are presented on histograms

Posterior analysis (model 2) [0-4 pts]

- were there any issues with the sampling? if there were what kind of ideas for mitigation were used
  1 pt - no issues with sampling

- are the samples from posterior predictive distribution analyzed
  1 pt - yes, there is histogram as justification with description

- are the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)
- **1 pt - data is quite consistent with posterior predictive samples, results visible on histogram**

- have parameter marginal distributions been analyzed (histograms of individual parameters plus summaries, are they diffuse or concentrated, what can we say about values)
- **1 pt – distribiutions are presented on histograms**

Model comparison [0-4 pts]

- Have models been compared using information criteria
  1 pt – yes, models have been compared

- Have result for WAIC been discussed (is there a clear winner, or is there an overlap, were there any warnings)
  1 pt - clear winner for WAIC is model 1, partial overlaps

- Have result for PSIS-LOO been discussed (is there a clear winner, or is there an overlap, were there any warnings)
  1 pt - clear winner for LOO is model 1, partial overlaps

- Whas the model comparison discussed? Do authors agree with information criteria? Why in your opinion one model better than another
  1 pt – yes, comparison was discussed