Assignment

November 2022

Descriptive statistics

Data	Min	Median	Mean	Max	Amount
TP	1.000	5.000	4.779	11.000	70
ME	1.000	4.500	4.479	9.000	24
LE	1.000	5.000	4.935	11.000	46

As shown in the table, the median is very similar between the different result for the total (TP) and the specific ones, ME (More Experienced) and LE (Less Experienced). The mean is also quite similar across the the total as well as the specific ones. However, there are a difference between the the max of ME compared to TP and LE. Me has a maximum of 9 while the other have a maximum of 11. Furthermore, we must consider that there is quite a larger amount of participants who is a part of the LE group. This can play a role in our result if we compare the two groups to each other.

Defence of the likelihood

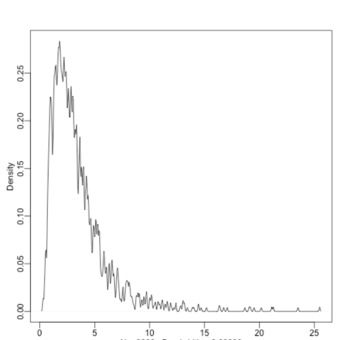

Figure 1: Gamma-Poisson model with the best outcome from model 1.

The graph above shows the best possible outcome with priors. Furthermore, model 2 is quite similar to model 1, hence model 3 instead of model 2 are shown here.

All models are Gamma-Poisson. In total there are four model, however one of the models are a null model. Hence, there are only three models to take into considerations.

Here are notations for model 1 and model 2:

$$\begin{gathered} tp \sim Gamma-Poisson(\lambda, \phi) \\ \log(\lambda) = \alpha + \beta_{[t_n um]} + \beta_{[c_n um]} \\ \alpha \sim Normal(0,1) \\ \beta_{[t_n um]} \sim Normal(0,1) \\ \beta_{[c_n um]} \sim Normal(0,1) \\ \phi \sim Exponential(1) \end{gathered}$$

$$\begin{gathered} tp \sim Gamma-Poisson(\lambda, \phi) \\ \log(\lambda) = \alpha_{[t_n um]} + \beta_{[c_n um]} \\ \alpha_{[t_n um]} \sim Normal(0,1) \\ \beta_{[c_n um]} \sim Normal(1,0.5) \\ \phi \sim Exponential(1) \end{gathered}$$

Figure 2: on the left model 2, and on the right model 1

The main reason that I choose to use Gamma-Poisson over any other model i.e. Poisson (mean and variance have to be roughly the same). If Poisson do not work, we will go to the next one, Gamma-Poisson, which is used to handle discrete distributions.

The differences in the models are the following: in the null model do I only use α and no β. In m1 is only one β (see img above) included for one parameter (NT). M2 has two β (see img above), one for NT and one for ME. M3 has no β, instead I combined values from ME and NT in a vector, afterwards I use a Normal distribution on them.

Since the mean and variance are not equal, we have to use the likelihood from Gamma-Poisson for the models. When executing the models, we can observe that some kind of pre-training is happening for the type count. We can also see that it only have positive numbers (integers). Hence that is exactly what the Poisson wants, however as mention before, the variance and mean are not the same, therefore Gamma-Poisson. When executing and plotting, we can compare them. There is a similarity between them (this include a coverage of the epistemological and the ontological perspective).

Prior for the final model

By testing the different models with the combinations of different priors and the comparing them against the data, model 2 had the best likelihood that matched. However, there is some differences between the them. To begin with the mean for the model is 4.881, while the counterpart have a mean of 4.779, approximately 0.1 in difference. The median for them are the same, a value of 5. Furthermore, the model has a much higher maximum than the raw data, 17 compared to 11. By testing these different priors, there is safe to state that there are a minimum difference by changing priors. However, by using priors, we do get a better result than just from the raw data.

Comparing models

The standard error for the models are the following:

Model	WAIC	SE	dSE	weight
null	609.6203	11.2076	3.5533	0.1856
1	608.3370	10.3877	1.7581	0.3526
2	607.7997	9.3093	NA	0.4613
3	622.2074	11.8397	5.7810	0.00034

The table summarises the Standard Error (SE) and the direct Standard Error (dSE). The null-model has the highest values, but it dose not have all parameters in it, that is why it is excluded. Therefore do we chose model 2 (m2) is the best candidate. It has the lowest dSE (NA which means that it is our best model) compared to the other models, it does also have the lowest SE. M2 does have the lowest WAIC as well compared to the other models.

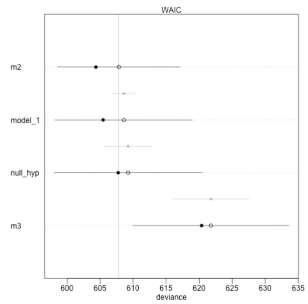

Figure 3: Plot of the WAIC with the different models, model 1, model 2, model 3, adn the null-model. Order of the models are: m2, m1, null & m3.

As shown in the plot, model 1, 2 & null are quite similar. However, m2 does have shorter lines than m1, and the null-model have a slightly longer line than m1. Furthermore, both model 1,2 & null are crossing over to m3. Since none of the models are separated, i.e. they are quite grouped, but shifted in either direction. If we look at the table above, we can see that m2 has the lowest dSE, hence m2 is our model choice.

DAG

The DAG is formed as a fork, where the category is affecting the the techniques and the tp. The reason why the DAG is designed as a fork is related to the data. By testing the LE against the ME, there is a higher mean and maximum amongst the LE rather then the ME. This tells us that the category have and effect on the techniques and the tp. This could have many reasons why, i.e. there could be a new technique the LE have learned in school, that the did not exist when the ME did study. Moreover, the amount of participants is also important to take in to consideration, there are 46 LE individuals and 24 ME individuals. Therefore, if the sample was more even distributed, there could might have been a case where the category did not impact the tp, however, that is not the case.

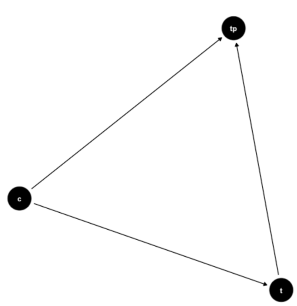

Figure 4: The DAG is a fork. The c stands for category, tp, and t for technique.

R hat

The $\hat{R}$ values for the models as well as the Bulk Effective Sample Size (BESS) and Tail Effective Sample Size (TESS) are the following for the models:

Model	$\hat{R}$	BESS	TESS
null	1.000131	2161.536	2029.742
1	1.003907	1907.131	1753.943
2	1.007832	1673.314	1793.672
3	1.035486	69.16797	1395.724

As shown in the table above, we can see that the null-model is the fastest one, however the null-model does not have any beta-parameters, hence it is faster. Moreover, model 1 is the fastest model when using both alpha and beta parameters. The $\hat{R}$ value is close to 1 for model 1,2 & null, but null has the lowest. Model 3 have a higher value. However, none of the models exceeds the value of 1.1, which means that non of them do converge. Moreover, we want a model with the lowest possible value closest to one. Hence I would choose model 2 since the null model does not take all parameters into account, and model 1 did not perform as well as model 2 when compared the two of them. However, model 1 does also have the highest BESS and TESS (excluding null-model), the higher value of these two, the better. However, model 2 did perform the best overall (see comparing models chapter).

Interpretation of the result

Firstly, when it comes the different participants with different levels of knowledge, the ontological assumption is that more experienced individuals will be better in their area of expertise than less experienced. However, by the epistemological part given the data, this assumption where the opposite. By doing statistics with the data, there is a possibility to show others that assumptions are not always correct.

```
              mean          sd         5.5%       94.5%
dif_t   -0.1300737  0.09646531   -0.291614  0.02232327
             mean           sd          5.5%        94.5%
dif_c   0.09603364  0.09838422  -0.06216378    0.2497935
```

There are strong correlation to the opposite of the observed statistics. This means that the perception of the statistics seems miss leading. There could also be something incorrect with the model.