1 Introduction

We have several data in our "*data.csv*". 35 different people have tried both old and new methods and depending of the person, this one is either less experienced, either he got more experience.

2 Likelihood

The likelihood used in this assignment is the Poisson likelihood. Indeed, it is adapted to count some events with low rate such as our true positive data As well, I tried with a Normal likelihood previously and I got bad results with *effective number* and the *caterpillar*. I didn't choose the Binomial distribution because the important point is the True Positive value to have an idea of how many faults the software engineer made. One thing I could do is categorize the error in 2 parts. If the subject has more than 10 true positives, he fails, either he got a success, but I didn't and focused on Poisson. You will see below that the final model seems coherent.

3 Models

I developed 4 different models, each ones run with map2stan. The more complex one (with the more variables) is the following :

$$
\begin{aligned}
\mathrm{N}_i &\sim \mathrm{Poisson}(\lambda_i)\\
\log(\lambda_i) &= \alpha + \beta_c * C_i + \beta_t * T_i + \beta_c t * C_i * T_i,\\
\alpha &\sim \mathrm{Normal}(0, 10)\\
\beta_c &\sim \mathrm{Normal}(0, 10)\\
\beta_t &\sim \mathrm{Normal}(0, 10)\\
\beta_c t &\sim \mathrm{Normal}(0, 10)
\end{aligned}
$$

For the other models I removed one term by one term. The second is : $\log(\lambda_i) = \alpha + \beta_c * C_i + \beta_t * T_i$
the third is: $\log(\lambda_i) = \alpha + \beta_t * T_i$
and the forth is : $\log(\lambda_i) = \alpha + \beta_c * C_i$

C is the category : 1 if the subject is More Experienced, 0 otherwise, T is the technique : 1 if the subject is using the New Technique, 0 otherwise. Beta, with X as indice, the corresponding letter of the parameter. Those notations are the same for the whole assignment.

The model *mstan1* is the model above : it is the more complex one. Model *mstan2* is the model without interactions between category and technique. *msan3 and mstan4* are models with either categories or either techniques.

After comparaison, the model 2 seems to be the best model to use as we can see with the high weight (0.69 is way more higher than 0.27). The pWAIC is not the lowest but the weight of mstan3 and mstan4 is so low (lower than 0.1) that having a pWAIC lower than the mstan1 is still a good indication.

Comparaison :

	WAIC	pWAIC	dWAIC	weight	SE	dSE
mstan2	323.7	5.5	0.0	0.69	22.02	NA
mstan1	325.5	6.9	1.8	0.27	21.72	1.79
mstan3	329.3	3.6	5.6	0.04	22.08	8.51
mstan4	418.3	3.3	94.6	0.00	15.71	14.76

$$\begin{aligned} \mathrm{N}_i &\sim \mathrm{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \alpha + \beta_c * C_i + \beta_t * T_i, \\ \alpha &\sim \mathrm{Normal}(0, 10) \\ \beta_c &\sim \mathrm{Normal}(0, 10) \\ \beta_t &\sim \mathrm{Normal}(0, 10) \end{aligned}$$

3.1 Priors

First, I used the model we choose and I have tried different prior with alpha only. I tried a Normal prior from 0 to 1;5;10;50.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
mstan1	322.8	5.0	0.0	0.32	22.05	NA
mstan3	323.2	5.3	0.4	0.26	21.90	0.40
mstan4	323.3	5.3	0.5	0.24	21.90	0.34
mstan2	323.9	5.6	1.1	0.19	22.03	0.61

The alpha with the Normal(0,1) seems to be the best options looking at pWAIC and weight. Let's try different prior with beta. I tried a Normal prior from 0 to 1;5;10, but I noticed the difference is tiny between each model. So, I have checked the caterpillar to have a better idea and the mstan3 is the best, the one with Norm(0,1) because others don't have very good hairy caterpillar.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
mstan2	323.1	5.2	0.0	0.40	21.92	NA
mstan1	323.6	5.4	0.4	0.32	21.93	0.36
mstan3	323.9	5.6	0.7	0.28	22.06	0.27

But when I So, this is my final model :

$$\begin{aligned} \mathrm{N}_i &\sim \mathrm{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \alpha + \beta_c * C_i + \beta_t * T_i, \\ \alpha &\sim \mathrm{Normal}(0, 1) \\ \beta_c &\sim \mathrm{Normal}(0, 1) \\ \beta_t &\sim \mathrm{Normal}(0, 1) \end{aligned}$$

Now we have to look at 3 things to know if the model is coherent :

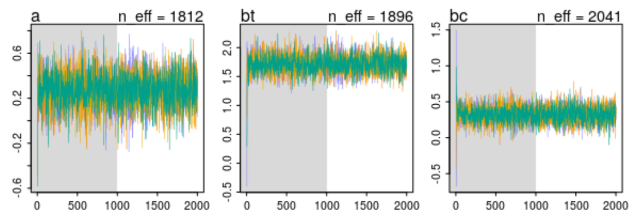

Figure 1: Caterpillar of our model

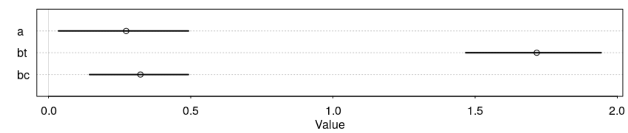

Figure 2: Plot the precis of the model

- Effective number higher than 0.1
- Hairy Caterpillar, as you can see in Figure 1.
- Rhat approaching 1 shows that the model build well. But it has to be lower than 1.1 because that strongly indicate that something is wrong.

Here is the result of the precis() function using 4 chains

	Mean	StdDev	lower 0.89	upper 0.89	n_eff	Rhat
a	0.27	0.15	0.07	0.53	1812	1
bt	1.72	0.15	1.48	1.95	1896	1
bc	0.32	0.11	0.15	0.51	2041	1

We have to notice that the effective number of samples are quite the same across the parameter. As you can see in Figure 2.

We can see that both effect of Bc, Bt and alpha are positive and quit precis (maximum 0.25 from the point). Alpha seems very close to zero. Bc is not as huge as the Bt but it is still positive. So we can notice the impact of both technique and experience on the True positive errors. Because Bt is larger than Bc, we can suppose that the Technique parameter has more effect.

So, it is highly recommended for a company to use the new technique because the effect seems clear and positive. About the experience, it is better to hire an experienced employee but considering a possible change of technique and the lower impact of experience, it is not mandatory to hire experience employee, it is depending of its needs.