

36663 Project Part 2

Mengran He

16 December, 2016

1. Response variable `reacTime`

1(a) In hierarchical linear models, the errors ϵ_i and random effects η_j are normally distributed with mean 0 and variance σ^2 and τ^2 respectively, because response variable $Y = X\beta + \epsilon$, the response variable also follows normal distribution in linear models. In this dataset, as shown in figure 1a, there are 8169 out of 8257 instances having reaction time less than 10 seconds, so the distribution of reaction time is very right skewed instead of bell shape. On the contrary, from figure 1b, the distribution of log of reaction time across 8257 instances is unimodal, which is more suitable to use as response variable for hierarchical linear model with normal errors and normal random effects.

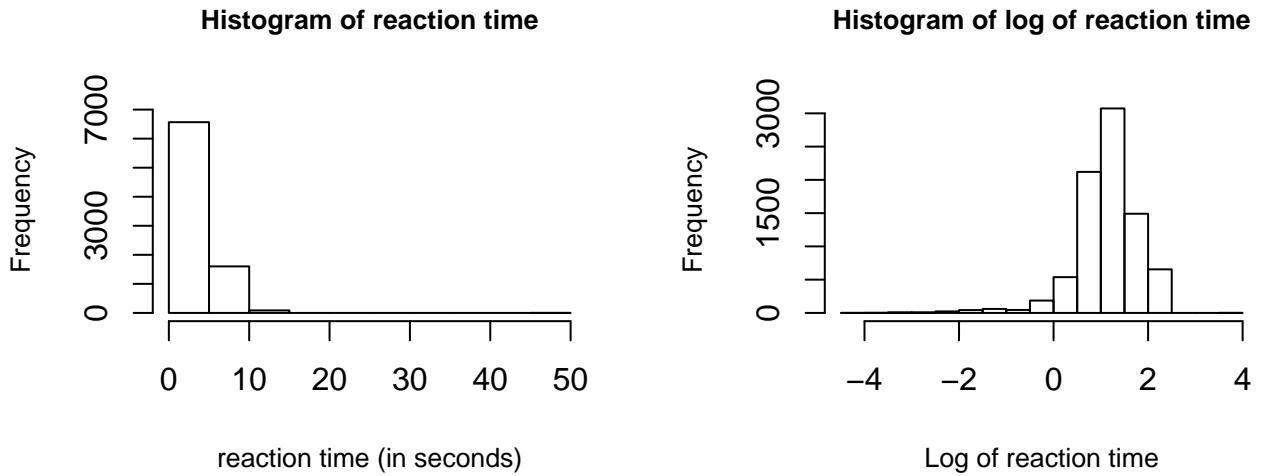


Figure 1: a) Histogram of reaction time, b) Log of reaction time

1(b) I If player's reaction time exceeds 10 seconds the time limit, his answer which is considered as wrong answer will be coded as `resp = 0` and `hitType= "Time Out!!"`. However, 250 instances whose reaction time is within time limit range were coded as wrong answer. So, the time limit doesn't strictly adhere to the data.

1(b) II Based on Derek Lomas' hypothesis, larger time limits might increase both success and engagement, so the time limit might had effect on player's reaction time. However, 250 instances didn't strictly adhere to the time limit, which might affect the accuracy of our linear model. For linear models, because the errors are generated by time limit and they can't be explained by the model, the variance of residuals may not be equal and there might be outliers in residual plots.

1(b) III According to Derek Lomas' paper, they tested 8 different time limits: 2, 3, 4, 5, 8, 10, 15 and 30 seconds (ship targets included an additional 5 seconds). In this dataset, the time limit is consistently 10 seconds, so we assumed that those 250 instances were from experiment with shorter time limits. In addition, for `reacTime` variable, the maximum value is 46.605 seconds with `hitType = "Time Out!!"`, so we also assumed that this instance was from other experiment with longer time limit 30 seconds. Therefore, for consistency reason, we should remove instances which have `hitType` equal "Time Out!!" but `reacTime` less than 10 seconds or equal to 46.605 seconds.

2. Ordinary linear regression for question effects

2(a) Fitting the linear model without the intercept means that each estimated coefficient is exactly the log of geometric mean of reaction time for each question. As we can see from table 1 that the estimated coefficients of all question types are statistically significant. Figure 2 shows the 95% confidence interval plot for 20 question effects on log of reaction time. From figure 2, we can see that question 0.16 and 0.8 have higher estimated coefficients, which means the reaction time of these two group is longer than other question groups. The geometric mean of reaction time is $3.2415 (=e^{1.17603})$ seconds

for question 0.16, the geometric mean of reaction time is $3.2093 (=e^{1.16606})$ seconds for question 0.8, while the geometric mean of reaction time is 2.7829 seconds for question 0.83.

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------|----------|------------|---------|----------|
| currentQuestion0.1 | 1.0941 | 0.0313 | 34.91 | 0.0000 |
| currentQuestion0.13 | 1.0987 | 0.0311 | 35.37 | 0.0000 |
| currentQuestion0.16 | 1.1760 | 0.0377 | 31.23 | 0.0000 |
| currentQuestion0.2 | 1.0890 | 0.0381 | 28.59 | 0.0000 |
| currentQuestion0.25 | 1.0557 | 0.0307 | 34.35 | 0.0000 |
| currentQuestion0.3 | 1.0903 | 0.0380 | 28.72 | 0.0000 |
| currentQuestion0.33 | 1.0771 | 0.0305 | 35.27 | 0.0000 |
| currentQuestion0.38 | 1.0739 | 0.0378 | 28.38 | 0.0000 |
| currentQuestion0.4 | 1.0867 | 0.0375 | 28.99 | 0.0000 |
| currentQuestion0.43 | 1.0822 | 0.0305 | 35.51 | 0.0000 |
| currentQuestion0.5 | 1.1185 | 0.0270 | 41.49 | 0.0000 |
| currentQuestion0.6 | 1.0884 | 0.0379 | 28.72 | 0.0000 |
| currentQuestion0.63 | 1.0765 | 0.0307 | 35.10 | 0.0000 |
| currentQuestion0.66 | 1.0976 | 0.0311 | 35.33 | 0.0000 |
| currentQuestion0.7 | 1.1070 | 0.0378 | 29.30 | 0.0000 |
| currentQuestion0.75 | 1.1017 | 0.0310 | 35.58 | 0.0000 |
| currentQuestion0.8 | 1.1661 | 0.0374 | 31.21 | 0.0000 |
| currentQuestion0.83 | 1.0235 | 0.0378 | 27.05 | 0.0000 |
| currentQuestion0.88 | 1.0749 | 0.0312 | 34.41 | 0.0000 |
| currentQuestion0.9 | 1.1072 | 0.0311 | 35.64 | 0.0000 |

Table 1: Summary of ordinary linear model

Confidence Interval for question effects

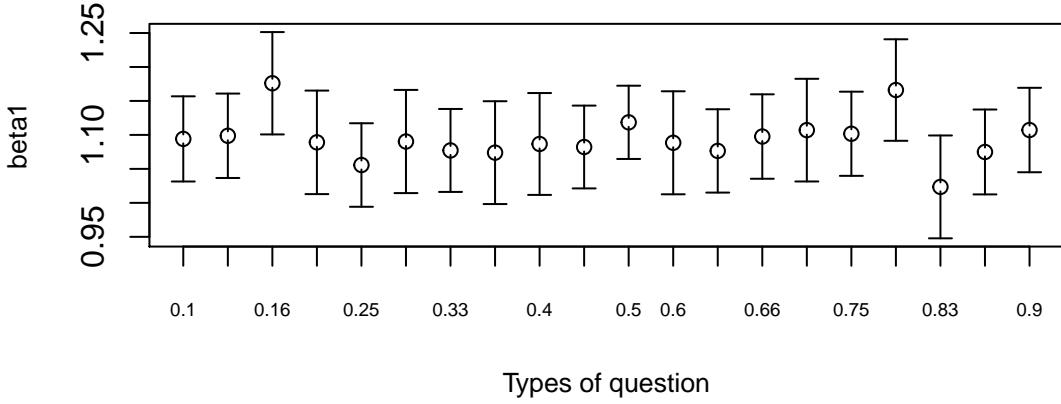


Figure 2: Plot of 95% confidence interval for each estimated coefficient in model 1

The diagnostic plots of `model11` in figure 3 shows that model 1 doesn't fit data very well. There are patterns in residuals vs fitted values plot and scale-location plot, and points on residuals' QQplot deviate from the normality assumption line a lot. So this linear model violates normaliy assumption and equal variance assumption.

2(b) The scatter plot of coefficients vs item easiness shows nearly horizontal positive regression line, indicating that the coefficients and item easiness has slightly positive relationship. The scatter plot of coefficients vs mean reaction time on each question shows a positively linear relationship as shown in figure 4a. The scatter plot of coefficients vs log of mean reaction time in figure 4b also suggests a positively linear relationship. The scatter plot of coefficients vs the target value of each question shows no linear relationship. By calculation, the correlation of $\hat{\beta}$ and mean reaction time, correlation of $\hat{\beta}$ and log of mean reaction time are both 0.83.

2(c) I The stepwise and forward model selection are used in automatic variable selection by comparing AIC of each model. The smallest model is `lm(lrt ~ currentQuestion - 1)` and largest model is `lm(lrt ~ resp + experimentName + levelName + isGuidesEnabled + currentQuestion + avgAccuracy + avgTime + bestTime + fireType + sound + currentStarCount)`. According to automated model selection, both methods chose the same

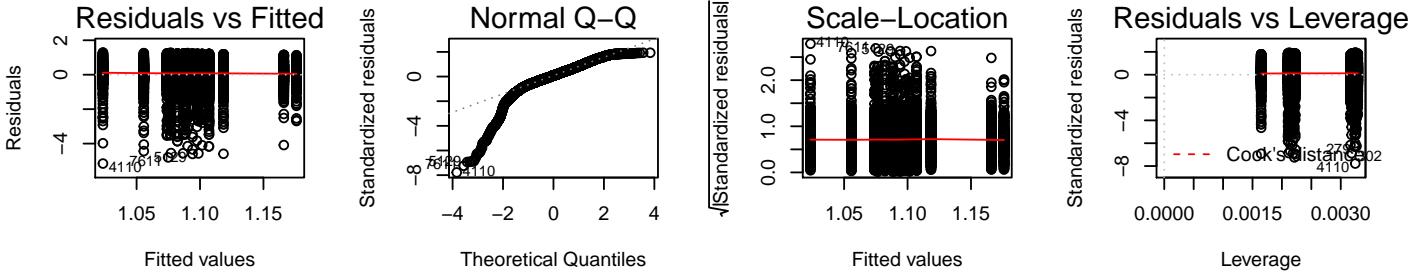


Figure 3: Diagnostic plots of model 1

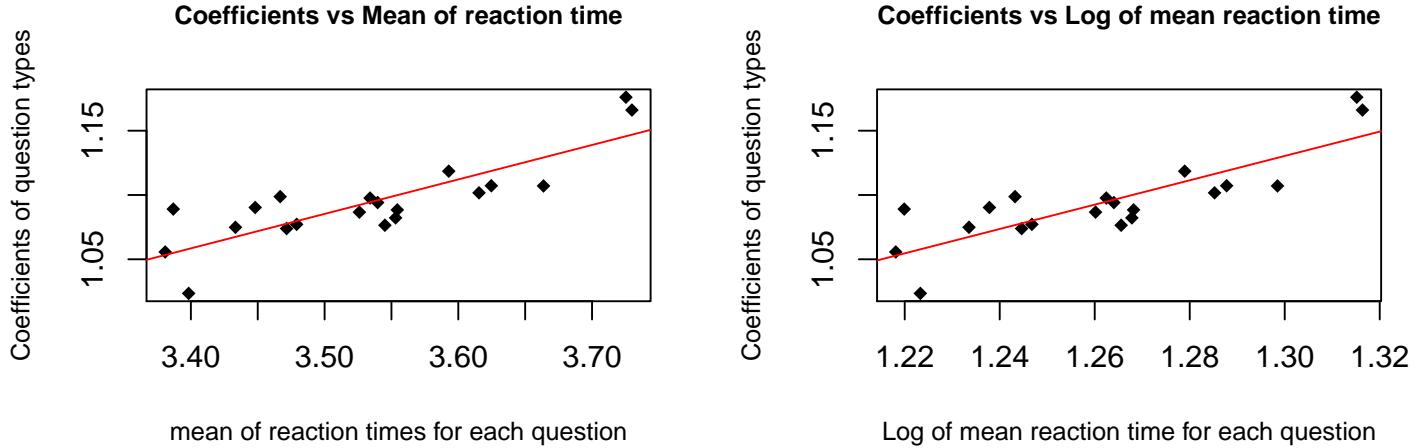


Figure 4: a)Coefficients vs Mean of reaction time, b) Coefficients vs Log of mean reaction time

model, which is `model1_best: lm(lrt ~ currentQuestion + avgTime + avgAccuracy + bestTime + levelName + currentStarCount + experimentName + fireType - 1)`.

2(c) II The summary of model `model1_best` is in appendix II. The estimated coefficients of 20 question types are all statistically significant as their P-values are less than 0.001. Question 0.16 has largest coefficient -0.7685, among 20 questions, which means the reaction time of question 0.16 is longer than reaction time of other questions. The smallest coefficient of question types is from question 0.83 where estimated coefficient = -0.9094, which means the reaction time of question 0.83 is shortest among 20 questions. `avgTime`, `avgAccuracy`, and `bestTime` are statistically significant, they have significant positive effects on the log of reaction time. We would expect the reaction time to increase 20%, 0.5% and 10.48% when there is one unit increase in `avgTime`, `avgAccuracy`, `bestTime` respectively. For `levelName` variable, there are 4 out of 7 levels not statistically significant, others have significantly negative effects on the log of reaction time. The coefficient of `currentStarCount` is positively related to the log of reaction time. There are 2 out of 15 levels of `experimentName` not statistically significant in this model, other levels have significantly positive effects on the log of reaction time. In addition, `fireType` = "CLICK" is also statistically significant, which has positive effect on the log of reaction time, so we would expect the reaction time to increase by 40% when player shifts from using keyboard to mouse click.

2(c) III `resp` is an indicator variable, 1 for correct answer, 0 for wrong answer, if `resp` becomes continuous from 0 to 1, they can no longer represent if the response is correct or not, so it doesn't make sense to explain its relationship with `lrt`. For variable `easiness` which measures the proportion of players attempting the item who got it right (`resp = 1`), while the response variable in the linear model is log of reactime time for each player, `easiness` can't be simply fitted in the ordinary linear model because variable length differ, but it can be fitted in multilevel model where `easiness` is in the second level. For variable `currentQuestion` where the numbers ("0.1", "0.13", "0.16", "0.2", "0.25", "0.3", "0.33", "0.38", "0.4", "0.43", "0.5", "0.6", "0.63", "0.66", "0.7", "0.75", "0.8", "0.83", "0.88", "0.9") represent the type of questions, if `currentQuestion` is continuous variable, the relationship between `currentQuestion` and `lrt` is explained by the change of the numbers. Nevertheless, the research doesn't indicate the meaning of name of questions, so the relationship between `lrt` and numeric `currentQuestion` doesn't make sense either.

3. Ordinary linear regression for player effects

3(a) After data cleaning in previous step, we have 407 unique player id. `model2: lm(lrt ~ SID - 1)` is linear regression model for player effect. There are 407 estimated coefficients representing the expected value of log of reaction time for 407 players. Figure 5 shows the histogram of coefficients of `model2`, over half of SID have coefficients between 1 and 1.5, which means the geometric mean of reaction time of over half of players is between 2.718 and 4.482 seconds. The maximum value of estimated coefficient is 2.347 corresponding to 10.45 seconds of reaction time. The minimum value of estimated coefficient is -1.146 corresponding to 0.32 seconds of reaction time. Among 407 SID player effects, nearly 90% of them are statistically significant. The diagnostic plot indicates that there are clear patterns in residuals vs fitted values plot and scale-location plot and points on the left tail of residuals' QQplot deviate from the straight line due to logarithms transformation.

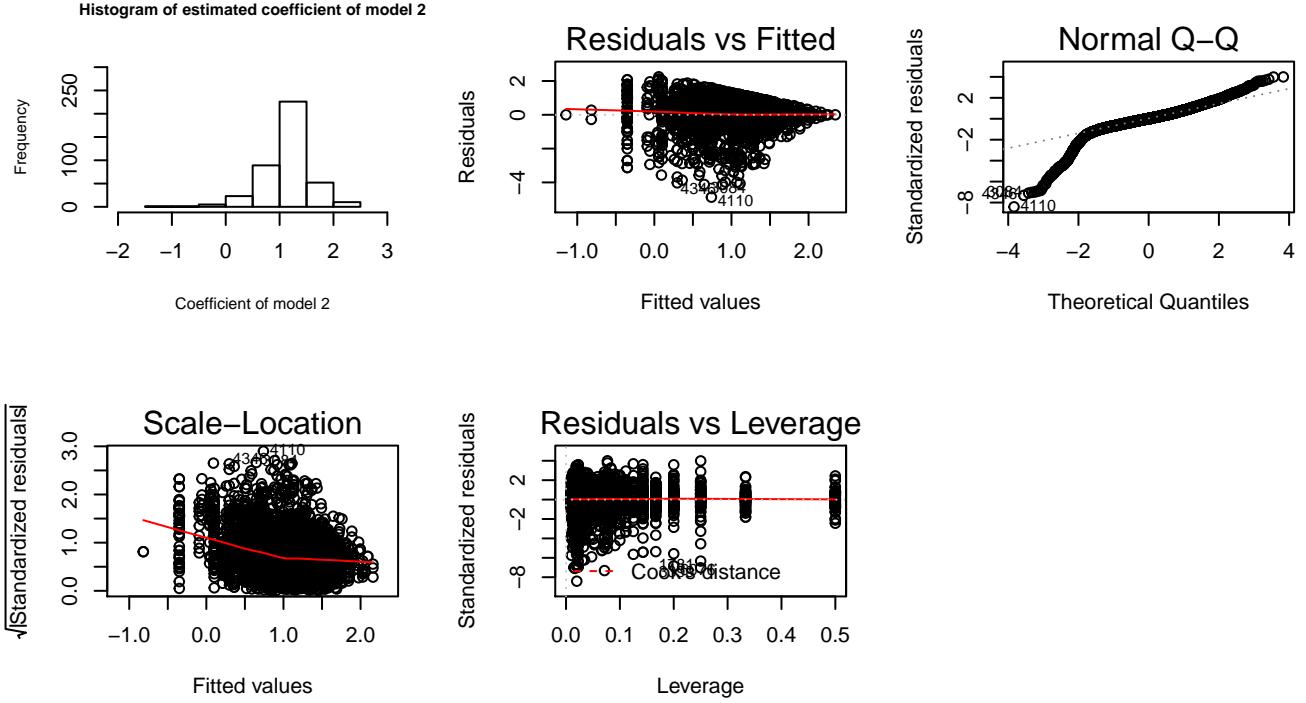


Figure 5: Histogram of estimated coefficient of model 2 and diagnostic plots of model 2

3b The scatter plot of coefficients against proportion correct doesn't identify any linear relationship, some points are lined up vertically. The scatter plot of coefficients against player random effects from model `glmer.resp <- glmer(resp ~ currentQuestion - 1 + (1|SID), family = binomial)` has positively linear relationship but slope is very flat. The scatter plot of coefficients vs mean reaction time of each player in figure 6a has inverted U shape, their correlation is 0.8879, so their relationship may not be linear. The scatter plot of coefficients vs log of mean reaction time for each play has perfectly positive linear trend, as shown in figure 6b, and their correlation is 0.9363842. Clearly, the estimated coefficients in `model2` are estimates of the log of mean reaction time for each player.

4. Mixed effects models

4(a) The mixed effects model `model_mix: lmer(lrt ~ currentQuestion + (1|sid) - 1)` took questions as fixed effects and players as random effects. Table 2 shows that the effects of all questions are statistically significant, among which question 0.16 has largest effects on the log of reaction time and question 0.83 has smallest effects on the log of reaction time, their effects on the log of reaction time are consistent with effects of previous ordinary linear model. Player would take longer reaction time on question 0.16 than doing other questions. From figure 7a, we can see that the values of coefficients mainly concentrate in range of 1.10 to 1.15. For mixed effect linear model, the residuals and random effects should follow normal distribution with mean 0. Figure 7b and 7c suggest that the QQplots of random effects and residuals of model `model_mix` support normality assumption, except that points on left tails of both QQplots deviate from the straight line. Figure 7d shows that there is a fan shape pattern in residuals vs fitted value plot, maybe it is because of time limit.

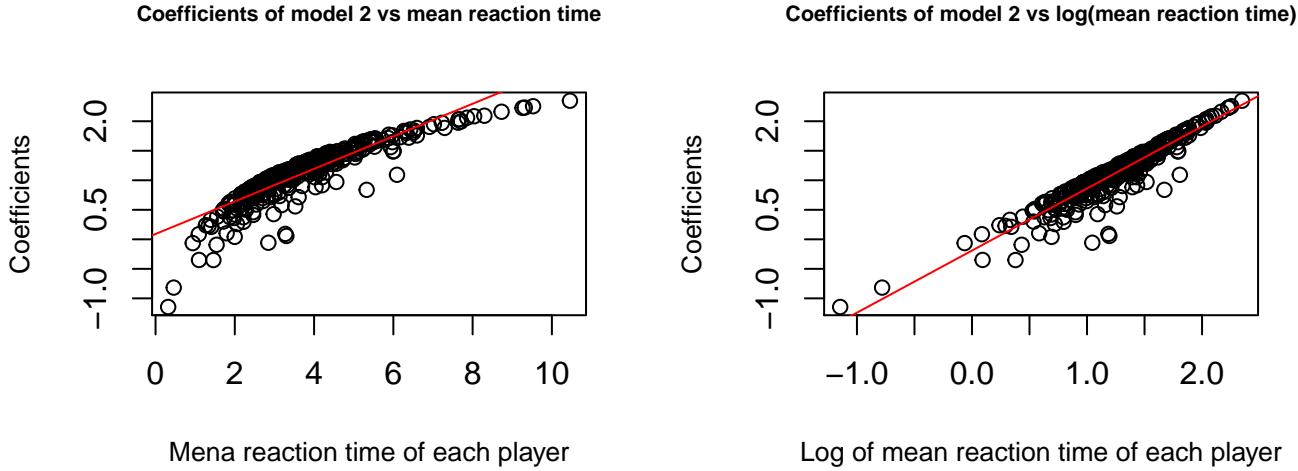


Figure 6: a) coefficients vs mean reaction time of player, b) coefficients vs log of mean reaction time of player

| | Estimate | Std. Error | t value |
|---------------------|----------|------------|---------|
| currentQuestion0.1 | 1.12 | 0.03 | 33.84 |
| currentQuestion0.13 | 1.11 | 0.03 | 33.74 |
| currentQuestion0.16 | 1.21 | 0.04 | 31.82 |
| currentQuestion0.2 | 1.12 | 0.04 | 29.25 |
| currentQuestion0.25 | 1.08 | 0.03 | 33.09 |
| currentQuestion0.3 | 1.12 | 0.04 | 29.13 |
| currentQuestion0.33 | 1.10 | 0.03 | 34.06 |
| currentQuestion0.38 | 1.10 | 0.04 | 28.74 |
| currentQuestion0.4 | 1.11 | 0.04 | 29.39 |
| currentQuestion0.43 | 1.10 | 0.03 | 33.98 |
| currentQuestion0.5 | 1.14 | 0.03 | 38.21 |
| currentQuestion0.6 | 1.12 | 0.04 | 29.25 |
| currentQuestion0.63 | 1.11 | 0.03 | 34.01 |
| currentQuestion0.66 | 1.11 | 0.03 | 33.80 |
| currentQuestion0.7 | 1.13 | 0.04 | 29.58 |
| currentQuestion0.75 | 1.13 | 0.03 | 34.24 |
| currentQuestion0.8 | 1.18 | 0.04 | 31.16 |
| currentQuestion0.83 | 1.05 | 0.04 | 27.55 |
| currentQuestion0.88 | 1.10 | 0.03 | 33.45 |
| currentQuestion0.9 | 1.12 | 0.03 | 34.11 |

Table 2: Summary of mixed effects model

4(b) I The largest model I considered was `full.lmer: lmer(lrt ~ resp + experimentName + levelName + isGuidesEnabled + currentQuestion + avgAccuracy + avgTime + bestTime + fireType + sound + currentStarCount + (1|SID) + (1|ip3) - 1)` and the smallest model was `lmer(lrt ~ currentQuestion + (1 | SID) - 1)`.

For automatic model selection, `fitLMER.fnc()` function in `LMERConvenienceFunctions` library was used to automate backwards selection of fixed effects and forward selection of random effects with BIC method. The full model without intercept is `full.lmer: lmer(lrt ~ resp + experimentName + levelName + isGuidesEnabled + currentQuestion + avgAccuracy + avgTime + bestTime + fireType + sound + currentStarCount + (1|SID) + (1|ip3) - 1)`, which considered SID and ip3 as random effects. Because ip2 and ip3 gave the same groups, so we only tried one of them. After `fitLMER.fnc()` model selection, the best model was `bic.best: lmer(lrt ~ isGuidesEnabled + avgAccuracy + avgTime + bestTime + currentStarCount + (1 | SID) + (1 | ip3) - 1)`, which kept both SID and ip3 as random effects. Then, `exactRLRT()` function in `RLRsim` library was used to test random effect of ip3. The result of Exact test shows RLRT = 1.1317, p-value = 0.1239, so we didn't reject the null: no random effect for ip3 in the model.

4(b) II Our research mainly focus on questions as fixed effects and players as random effect, we should include `currentQuestion` into the model we just chosen, so the final model is `final.lmer: lmer(lrt ~ currentQuestion + isGuidesEnabled + avgAccuracy + avgTime + bestTime + currentStarCount + (1 | sid) - 1)`. The summary of fixed effects of `final.lmer` is shown in appendix II, all of fixed effects are statistically significant. In particular, the

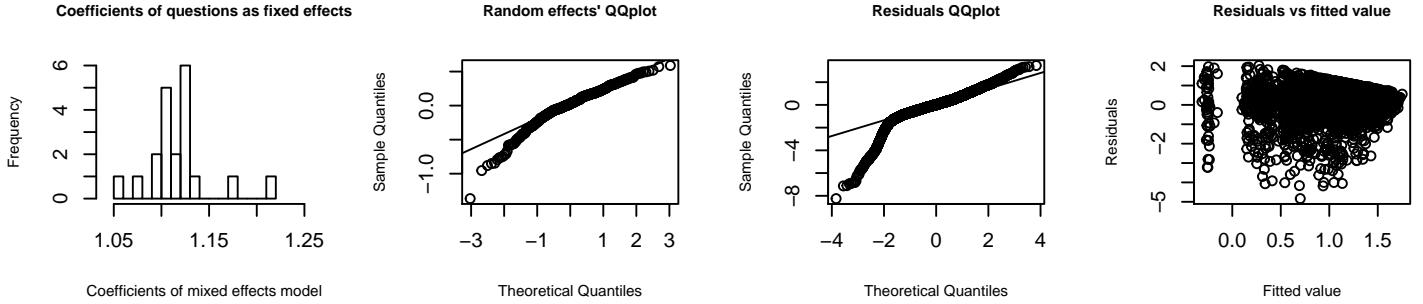


Figure 7: a) Histogram of coefficients of questions as fixed effects, b) QQplot of random effects, c) QQplot of residuals

reaction time is expected to increase 19.13% for one unit increase in `avgTime`, and the reaction time is expected to increase 5.05% if guide is enabled. In figure 8a and 8b, the QQplots of residuals and random effects indicate that the points on lower left tail deviating from straight line, maybe because of logarithms transformation on response variable. We still found pattern in residuals vs fitted value plot in figure 8c.

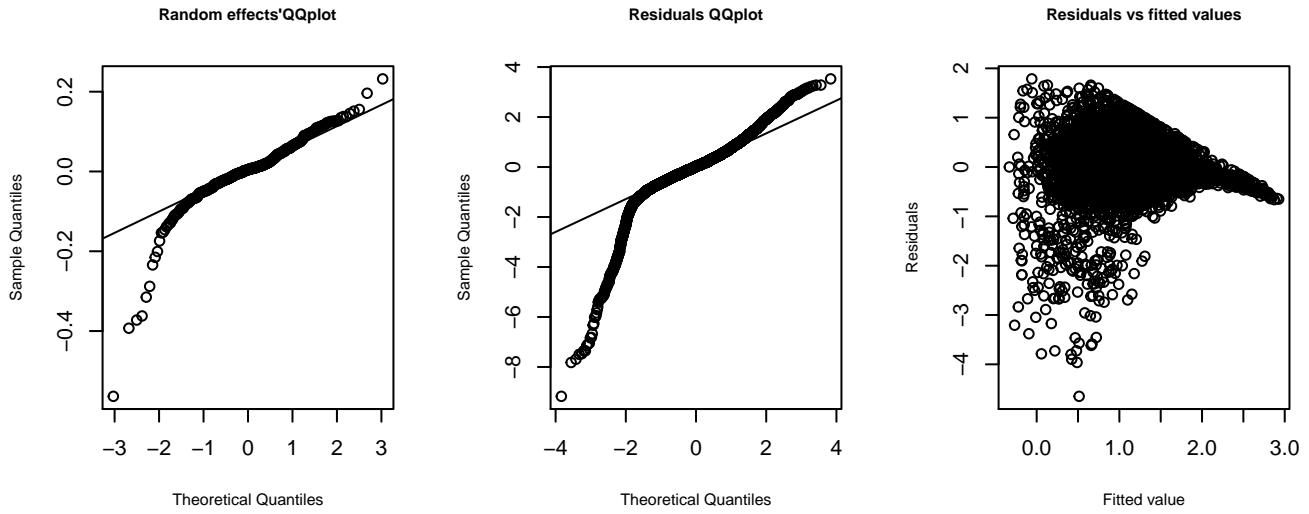


Figure 8: a) Random effects QQplot, b)Residuals QQplot, c)Residuals vs fitted values

5. Models combining reaction time and correctness of response.

5(a) As shown in figure 9a, the item easiness and the mean reaction time of each question has some positive linear relationship based on the instances collected, it suggests that the easier the question is, the longer the reaction time. According to Derek Lomas's paper, they measured player engagement by duration of total time played (sum of reaction times across estimates), and they concluded that if the game is easy, player is more willing to engage in the time which results in longer reaction time. Figure 9b suggests that the reaction time of player doesn't depend on the correct rate of player. There are lots of points lined up at proportion correct = 0, it suggests that proportion correct rate of player is 0, the average reaction time of player varies from 0 to more than 10 seconds. When player's proportion correct is 0.5, the average reaction time varies from 0 to about 6 seconds. Moreover, model `glmer.reac: glmer(resp ~ reacTime - 1 + (1|sid))` suggests that for 10% increase in reaction time, the probability of correct response is expected to decrease by 1.4%. In figure 9c the plot of predicted probability of correct response vs reaction time, we can see that given the increase of reaction time, the probability of correct response slightly decreases but doesn't change much. There is a spike at 10 seconds because of time limit. Therefore, we concluded that when question gets harder to answer correctly, the reaction time of player doesn't go up very much, but if the question gets easier, the reaction time gets longer due to player's engagement to the game.

5(b) I We fitted a linear regression model `linden.lm: lm(lrt ~ lp)`, as shown in appendix II, the summary of `linden.lm`. The intercept and the predictor `lp` are statistically significant as the p-values are less than 0.001. The coefficient of `lp` is positive, suggesting that `lp` and `lrt` has positively linear relationship. In figure 10, the diagnostic plots suggest that there are many negative residuals and some outliers in the model, but residuals are independent from fitted values in general.

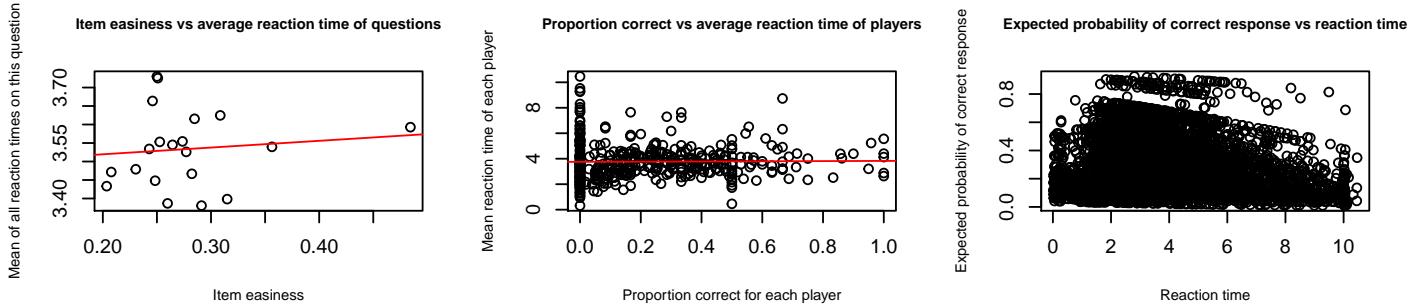


Figure 9: a) Item easiness vs average reaction time of questions, b) Proportion correct vs average reaction time of players, c) Expected probability of correct response vs reaction time

Residuals QQplot shows long negative tails due to logarithms transformation, the scale-location plot support the equal variance assumption. However, from the scatter plot of `lrt` vs `lp` we can see that there are lots of variability, so `lp` alone is not good enough for predicting the log of reaction time. The indication of positive relationship agrees with my previous statement, if the correct response is high, players may be more motivated to play more game, contributing more time to the questions, which may in turn increases reaction time.

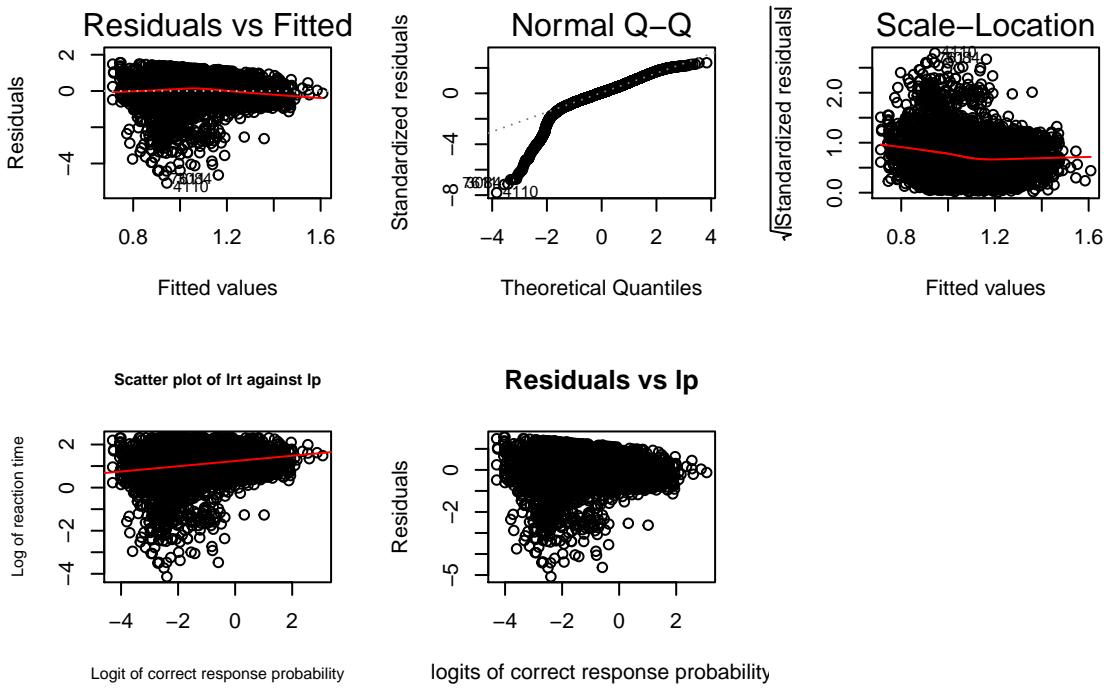


Figure 10: diagnostic plots of model `linden.lm`

5(b) II We applied stepwise and backward automatic model selection on full model `linden.full`: `lm(lrt ~ lp + currentQuestion + avgAccuracy + avgTime + bestTime)` by measuring AIC, and the model selection methods produce the same model `linden.aic`: `lrt ~ lm(lp + currentQuestion + avgAccuracy + avgTime + bestTime)`. Figure 11 shows the diagnostic plots of model `linden.aic`, by comparing with initial model `linden.lm`, we can see that the residuals are independent from fitted values and covariates in `linden.aic`, and residuals variance tend to be equal. Although the residual QQplot still has heavy tail on the lower left, it is due to log transformation where overfitting might exists, but in general `linden.aic` is better than `linden.lm`. The summary of `linden.aic` shown in appendix II indicates that covariates `lp`, `avgAccuracy`, `avgTime`, `bestTime` are statistically significant. For one unit increase in logit of success probability `lp`, `lrt` is expected to increase by 2.64%, for one unit increase in `avgAccuracy`, `lrt` is expected to increase by 0.56%, for one unit increase in `avgTime`, `lrt` is expected to increase by 20.5%, for one unit increase in `bestTime`, `lrt` is expected to increase by 8.62%. For `currentQuestion`, there are 3 out of 20 questions are statistically significant.

5(c) I

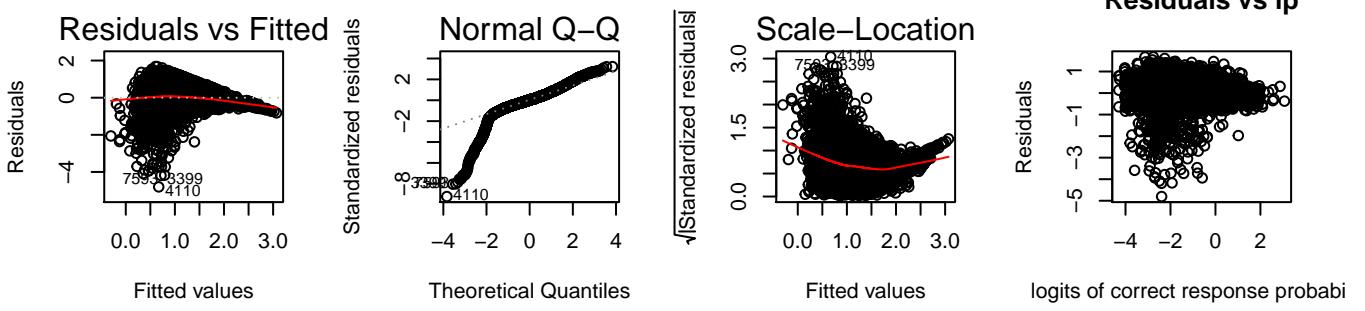


Figure 11: Diagnostic plots of model linden.aic

$$\text{Level 1: } lrt = \rho p_{j[i]} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

$$lp_j = \beta_j currentQuestion_j + \alpha_{l[j]} player_j$$

$$\text{Level 2: } \alpha_l = \beta_0 + \eta_l, \eta_l \sim N(0, \tau^2)$$

5(c) II A For model joint.mcmc.0, τ^2 has Rhat = 47.1, indicating that the Markov Chain hasn't converged to its stationary distribution. Also, from the ACF plot, there are lots of spikes along the lag, indicating the autocorrelation exists. Same problems exist in β_0 and β_j , and for α_l , the autocorrelation problem is most serious. Because joint.mcmc.0 doesn't produce convergent stationary distribution, it is not reliable to account them as posterior distribution. For model joint.mcmc.1, the Markov Chain converges to its stationary distribution, as the parameters don't have autocorrelation, and three chains for each estimate are on top of each other, Rhat for all estimates are close to 1, which is good indication of convergence, so the estimates from joint.mcmc.1 are useful and reliable.

5(c) II B The difference is setup of initial values. For joint.mcmc.0, initial values of bb, rho, and b0 have mean 0 and variance 1, and initial values for τ and σ are chosen between 0 to 1 from uniform distribution. For joint.mcmc.1, the estimates' initial values are chosen from the posterior of joint.mcmc.0. The initial values are the initial state of Markov Chain and they will drift to the posterior distribution after many iterations. If the initial values are so different from the posterior distribution, we need more to make Markov Chains converge. joint.mcmc.1 has initial values closer to posterior, making Markov Chains easily converging to stationary distribution.

5(c) III The summary of estimates of model joint.mcmc.1 are in table 3, where the fixed effects and random effects for generating 1p are included, the mean estimated coefficient of 1p is -1.2608 and the mean estimaed intercept is -0.9406, and they are both statistically significant in this model. For one unit increase in the logit of success probability 1p, the log of reaction time lrt is expected to decrease by 1.26, indicating the negative linear relationship, so if the probability of answering correctly increases, the reaction time tends to decrease. On the other hand, model in part b(i) linden.lm suggests a positive relationship between lrt and 1p with estimated coefficient of 1p = 0.1218, intercept = 1.2351. The conclusion in these two models are different.

| | mean | sd | twoSE | significance |
|-----------|---------|--------|--------|--------------|
| alpha[1] | -0.6191 | 0.0597 | 0.1194 | 1 |
| alpha[2] | -1.0963 | 0.1766 | 0.3532 | 1 |
| alpha[3] | -0.8456 | 0.1226 | 0.2452 | 1 |
| alpha[4] | -0.9517 | 0.1578 | 0.3155 | 1 |
| alpha[5] | -0.6951 | 0.0806 | 0.1613 | 1 |
| alpha[6] | -0.7528 | 0.1622 | 0.3244 | 1 |
| alpha[7] | -1.0220 | 0.1673 | 0.3346 | 1 |
| alpha[8] | -0.8287 | 0.1530 | 0.3060 | 1 |
| alpha[9] | -1.1671 | 0.0845 | 0.1690 | 1 |
| alpha[10] | -1.0465 | 0.1515 | 0.3031 | 1 |
| b[1] | 0.0053 | 0.0215 | 0.0430 | 0 |
| b[2] | 0.0033 | 0.0222 | 0.0445 | 0 |
| b[3] | -0.0623 | 0.0265 | 0.0530 | 1 |
| b[4] | -0.0095 | 0.0265 | 0.0531 | 0 |
| b[5] | 0.0327 | 0.0219 | 0.0438 | 0 |
| b[6] | 0.0075 | 0.0265 | 0.0529 | 0 |
| b[7] | -0.0003 | 0.0210 | 0.0420 | 0 |
| b[8] | 0.0054 | 0.0264 | 0.0527 | 0 |
| b[9] | 0.0107 | 0.0259 | 0.0518 | 0 |
| b[10] | 0.0080 | 0.0214 | 0.0427 | 0 |
| b0 | -0.9406 | 0.0267 | 0.0533 | 1 |
| rho | -1.2608 | 0.0313 | 0.0625 | 1 |
| sig | 0.6169 | 0.0050 | 0.0100 | 1 |
| tau | 0.2318 | 0.0136 | 0.0272 | 1 |

Table 3: Summary of joint.mcmc.1 (significance = 1 if absolute value of mean is greater than 2*SE)

Appendix I

1. Lomas, Derek, Kishan Patel, Jodi L. Forlizzi, and Kenneth R. Koedinger. “Optimizing Challenge in an Educational Game Using Large-scale Design Experiments.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13 (2013): n. pag. Web.
2. Gelman, Andrew, and Jennifer Hill. Data Analysis Using Regression and Multilevel/hierarchical Models. Cambridge: Cambridge UP, 2007.

Appendix II

| | Estimate | Std. Error | t value |
|---------------------|----------|------------|---------|
| currentQuestion0.1 | -0.35 | 0.05 | -6.37 |
| currentQuestion0.13 | -0.36 | 0.05 | -6.48 |
| currentQuestion0.16 | -0.26 | 0.06 | -4.49 |
| currentQuestion0.2 | -0.32 | 0.06 | -5.46 |
| currentQuestion0.25 | -0.37 | 0.06 | -6.74 |
| currentQuestion0.3 | -0.34 | 0.06 | -5.88 |
| currentQuestion0.33 | -0.37 | 0.05 | -6.68 |
| currentQuestion0.38 | -0.35 | 0.06 | -6.01 |
| currentQuestion0.4 | -0.36 | 0.06 | -6.20 |
| currentQuestion0.43 | -0.37 | 0.06 | -6.77 |
| currentQuestion0.5 | -0.39 | 0.05 | -7.11 |
| currentQuestion0.6 | -0.34 | 0.06 | -5.93 |
| currentQuestion0.63 | -0.37 | 0.06 | -6.64 |
| currentQuestion0.66 | -0.35 | 0.06 | -6.27 |
| currentQuestion0.7 | -0.33 | 0.06 | -5.76 |
| currentQuestion0.75 | -0.36 | 0.06 | -6.55 |
| currentQuestion0.8 | -0.28 | 0.06 | -4.85 |
| currentQuestion0.83 | -0.40 | 0.06 | -6.93 |
| currentQuestion0.88 | -0.37 | 0.05 | -6.77 |
| currentQuestion0.9 | -0.34 | 0.05 | -6.21 |
| isGuidesEnabledTRUE | 0.05 | 0.01 | 3.64 |
| avgAccuracy | 0.00 | 0.00 | 10.21 |
| avgTime | 0.19 | 0.01 | 25.15 |
| bestTime | 0.10 | 0.01 | 15.32 |
| currentStarCount | 0.02 | 0.00 | 6.74 |

Table 4: Summary of fixed in final.lmer model in Q4 part b

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 1.24 | 0.01 | 116.42 | 0.00 |
| lp | 0.12 | 0.01 | 18.31 | 0.00 |

Table 5: Summary of linear model lrt lp in Q5 part b(i)

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------|----------|------------|---------|----------|
| (Intercept) | -0.34 | 0.06 | -6.04 | 0.00 |
| lp | 0.03 | 0.01 | 3.74 | 0.00 |
| currentQuestion0.13 | 0.01 | 0.03 | 0.15 | 0.88 |
| currentQuestion0.16 | 0.11 | 0.04 | 2.89 | 0.00 |
| currentQuestion0.2 | 0.06 | 0.04 | 1.53 | 0.13 |
| currentQuestion0.25 | -0.02 | 0.03 | -0.59 | 0.56 |
| currentQuestion0.3 | 0.04 | 0.04 | 1.07 | 0.28 |
| currentQuestion0.33 | -0.01 | 0.03 | -0.25 | 0.80 |
| currentQuestion0.38 | 0.03 | 0.04 | 0.88 | 0.38 |
| currentQuestion0.4 | 0.01 | 0.04 | 0.33 | 0.74 |
| currentQuestion0.43 | -0.01 | 0.03 | -0.39 | 0.70 |
| currentQuestion0.5 | -0.07 | 0.03 | -2.08 | 0.04 |
| currentQuestion0.6 | 0.03 | 0.04 | 0.89 | 0.38 |
| currentQuestion0.63 | -0.01 | 0.03 | -0.40 | 0.69 |
| currentQuestion0.66 | 0.01 | 0.04 | 0.40 | 0.69 |
| currentQuestion0.7 | 0.06 | 0.04 | 1.45 | 0.15 |
| currentQuestion0.75 | -0.01 | 0.03 | -0.34 | 0.73 |
| currentQuestion0.8 | 0.11 | 0.04 | 2.74 | 0.01 |
| currentQuestion0.83 | -0.04 | 0.04 | -0.94 | 0.35 |
| currentQuestion0.88 | -0.01 | 0.04 | -0.22 | 0.82 |
| currentQuestion0.9 | 0.02 | 0.03 | 0.52 | 0.60 |
| avgAccuracy | 0.01 | 0.00 | 11.98 | 0.00 |
| avgTime | 0.21 | 0.01 | 30.43 | 0.00 |
| bestTime | 0.09 | 0.01 | 14.41 | 0.00 |

Table 6: Summary of linear model in Q5 part b(ii)

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------------------|----------|------------|---------|----------|
| currentQuestion0.1 | -0.8273 | 0.1334 | -6.20 | 0.0000 |
| currentQuestion0.13 | -0.8427 | 0.1339 | -6.29 | 0.0000 |
| currentQuestion0.16 | -0.7685 | 0.1345 | -5.72 | 0.0000 |
| currentQuestion0.2 | -0.8219 | 0.1346 | -6.11 | 0.0000 |
| currentQuestion0.25 | -0.8611 | 0.1338 | -6.44 | 0.0000 |
| currentQuestion0.3 | -0.8423 | 0.1340 | -6.28 | 0.0000 |
| currentQuestion0.33 | -0.8487 | 0.1333 | -6.37 | 0.0000 |
| currentQuestion0.38 | -0.8534 | 0.1347 | -6.34 | 0.0000 |
| currentQuestion0.4 | -0.8653 | 0.1344 | -6.44 | 0.0000 |
| currentQuestion0.43 | -0.8603 | 0.1331 | -6.46 | 0.0000 |
| currentQuestion0.5 | -0.8630 | 0.1336 | -6.46 | 0.0000 |
| currentQuestion0.6 | -0.8457 | 0.1341 | -6.31 | 0.0000 |
| currentQuestion0.63 | -0.8501 | 0.1338 | -6.35 | 0.0000 |
| currentQuestion0.66 | -0.8357 | 0.1338 | -6.24 | 0.0000 |
| currentQuestion0.7 | -0.8318 | 0.1341 | -6.20 | 0.0000 |
| currentQuestion0.75 | -0.8470 | 0.1337 | -6.33 | 0.0000 |
| currentQuestion0.8 | -0.7806 | 0.1344 | -5.81 | 0.0000 |
| currentQuestion0.83 | -0.9094 | 0.1348 | -6.75 | 0.0000 |
| currentQuestion0.88 | -0.8540 | 0.1334 | -6.40 | 0.0000 |
| currentQuestion0.9 | -0.8251 | 0.1339 | -6.16 | 0.0000 |
| avgTime | 0.2010 | 0.0069 | 29.18 | 0.0000 |
| avgAccuracy | 0.0049 | 0.0005 | 10.42 | 0.0000 |
| bestTime | 0.1049 | 0.0064 | 16.43 | 0.0000 |
| levelNameDecimal_3rdsG | -0.0154 | 0.0439 | -0.35 | 0.7254 |
| levelNameDecimal_4thsG | -0.0136 | 0.0427 | -0.32 | 0.7494 |
| levelNameDecimal_MidG | 0.0143 | 0.0411 | 0.35 | 0.7279 |
| levelNameDecimal_NoG | -0.0131 | 0.0304 | -0.43 | 0.6661 |
| levelNameDecimal_pretestA | -0.1231 | 0.0348 | -3.54 | 0.0004 |
| levelNameDecimal_pretestB | -0.1419 | 0.0349 | -4.06 | 0.0000 |
| currentStarCount | 0.0164 | 0.0025 | 6.61 | 0.0000 |
| experimentNameDecimal_10thsG_AB | 0.1236 | 0.0334 | 3.70 | 0.0002 |
| experimentNameDecimal_10thsG_BA | 0.1365 | 0.0373 | 3.66 | 0.0003 |
| experimentNameDecimal_3rdsG_AB | 0.0319 | 0.0555 | 0.58 | 0.5652 |
| experimentNameDecimal_3rdsG_BA | 0.1375 | 0.0436 | 3.15 | 0.0016 |
| experimentNameDecimal_3rdsG_AB | 0.0969 | 0.0459 | 2.11 | 0.0347 |
| experimentNameDecimal_4thsG_AB | 0.1096 | 0.0523 | 2.10 | 0.0359 |
| experimentNameDecimal_4thsG_BA | 0.1616 | 0.0428 | 3.78 | 0.0002 |
| experimentNameDecimal_4thsG_BA | 0.1102 | 0.0426 | 2.59 | 0.0097 |
| experimentNameDecimal_midG_AB | 0.0437 | 0.0487 | 0.90 | 0.3688 |
| experimentNameDecimal_MidG_AB | 0.1775 | 0.0430 | 4.13 | 0.0000 |
| experimentNameDecimal_MidG_BA | 0.1724 | 0.0406 | 4.25 | 0.0000 |
| experimentNameDecimal_NoG_AB | 0.1311 | 0.0393 | 3.34 | 0.0008 |
| experimentNameDecimal_NoG_AB | 0.1316 | 0.0412 | 3.19 | 0.0014 |
| experimentNameDecimal_NoG_BA | 0.0858 | 0.0395 | 2.17 | 0.0298 |
| fireTypeCLICK | 0.4027 | 0.1287 | 3.13 | 0.0018 |

Table 7: Summary of best model after automated model selection in Q2 part c