

Project Milestone 2 Report | TEAM 23

Sai Kiran Peruru (UH ID : 2288947) | Mohammed Adil Ahmed (UH ID : 2304590)

2024-05-02

Contents

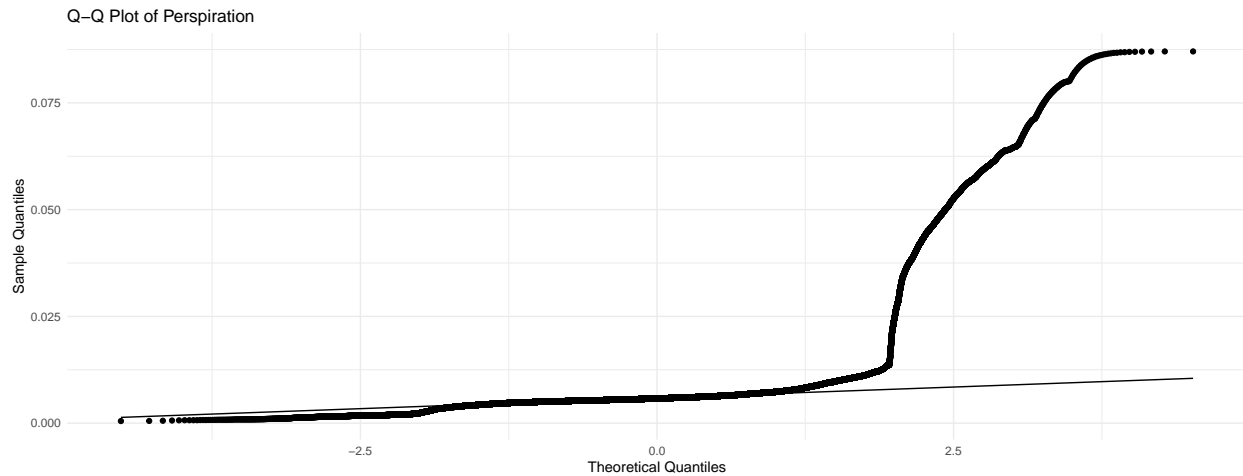
Data Processing	3
Checking Perspiration Data	3
QQPlot analysis of Perspiration Data	3
BoxPlot analysis of Perspiration Data	3
Baseline Data Processing	4
Exam data Processing	4
Structured data for final version of Exam data	5
Collinearity Check for Quantitative and Categorical Predictors	6
Step 1	8
Full Model 1 - Normalized PP_log with Participant centered Random Effect	8
Optimal Model by “Backward” Elimination	10
Normalized PP_log Optimal Model by BE	10
Comparison of Full and Optimal Models	12
Optimal Model Plots	13
Estimated Effects Plot	13
Random Effects Plot	14
Predictors Plot	15
Step 2	17
Full Model 2 - Normalized Hr E4 Perfect with Participant centered Random Effect	17
Optimal Model by Backward Elimination	19
Normalized HRE4 Perfect Optimal Model by BE	19
Comparison of Full and Optimal Models	21
Final Model Plots	22
Estimated Effects Plot	22
Random Effects Plot	23
Predictors plot	24
Step 3	26
Full Model 3 - Normalized HRV IBI with Participant centered Random Effect	26
Optimal Model by Backward Elimination	28
Normalized HRV.IBI Optimal Model by BE	28
Comparison of Full and Optimal Models	30
Final Model Plots	31
Estimated Effects Plot	31
Random Effects Plot	32
Predictors plot	33
Comparison Plot for Three Stress Variables	34

Step 4	35
Full Model	35
Manual Optimization:	37
Removing sPPlogNorm	37
Removing sHRE4PerfNorm	38
Removing sHRVIBINorm	40
Removing QType	41
Final Model	43
Comparison of Full and Optimal Models	44
Random Effects Plot	46
Predictors Plot	47
Final Conclusions:	48

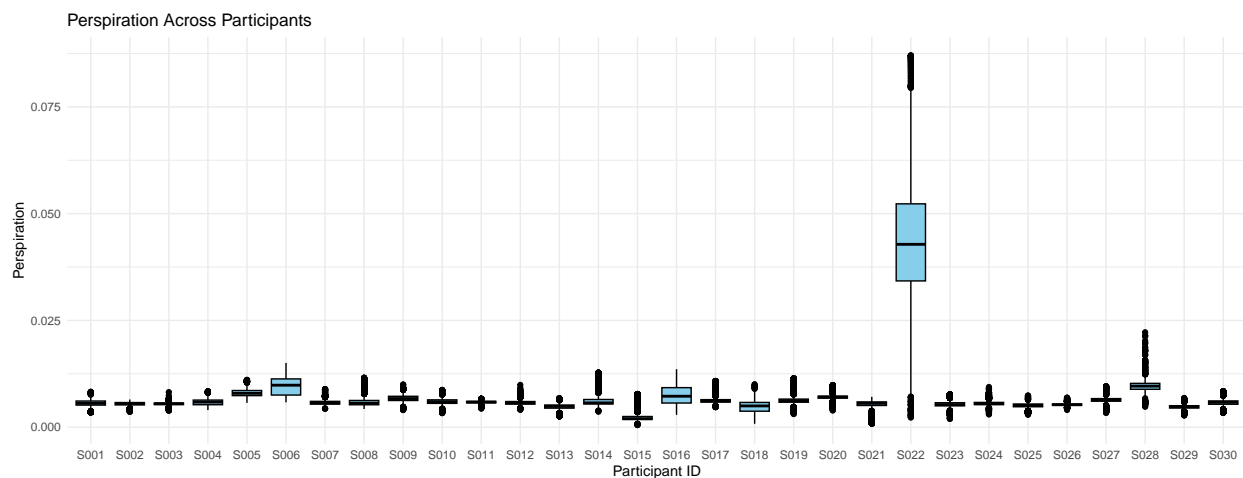
Data Processing

Checking Perspiration Data

QQPlot analysis of Perspiration Data



BoxPlot analysis of Perspiration Data



The QQ plot shows the quantiles of the perinasal perspiration data against the expected quantiles of a normal distribution. The S-shaped curve in the plot suggests that the distribution of perspiration is not normal. The lower and upper tails of the data curve away significantly from the line that represents a normal distribution. This indicates heavy tails higher frequencies of extremely low and high values which is typical for data that are not symmetrical and may contain outliers.

From Boxplot we can say most participants show low variability in their perspiration levels, as indicated by the tightly packed boxes. However, certain participants exhibit higher median perspiration levels and a broader range of values, as indicated by the larger boxes and whiskers. Participant S022 showed a huge variation indicating there is some abnormality in data. So S022 is removed and further investigated, which yielded a better results.

Baseline Data Processing

```
## 'data.frame': 8752 obs. of 9 variables:
## $ ParticipantID: Factor w/ 30 levels "S001","S002",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Perspiration : num 0.00595 0.0059 0.0059 0.00593 0.00593 ...
## $ HR.E4 : num 87.8 87.7 87.7 87.6 87.5 ...
## $ HR.AW : int NA NA NA NA NA NA 88 NA NA NA ...
## $ HRV.IBI : num NA NA 0.734 0.719 0.703 ...
## $ Breathing : num NA NA NA NA NA NA NA NA NA NA ...
## $ Time : int 0 1 2 3 4 5 6 7 8 9 ...
## $ pp_log : num -5.12 -5.13 -5.13 -5.13 -5.13 ...
## $ HR.E4_perfect: num NA NA NA NA NA NA 87.5 NA NA NA ...
```

Baseline data processing involves using baseline measurements as a reference point to assess changes or effects during an experimental condition. Subtracting baseline values from during-exam values normalizes the data, allowing for a clearer understanding of the physiological impact of the exam itself, independent of individual baseline differences. By subtracting baseline data, you can enhance the sensitivity of your analysis to detect changes due to the exam.

Exam data Processing

The Exam data processing helps in analyzing the physiological and performance responses of participants in an experimental math exam. The process that we followed here:

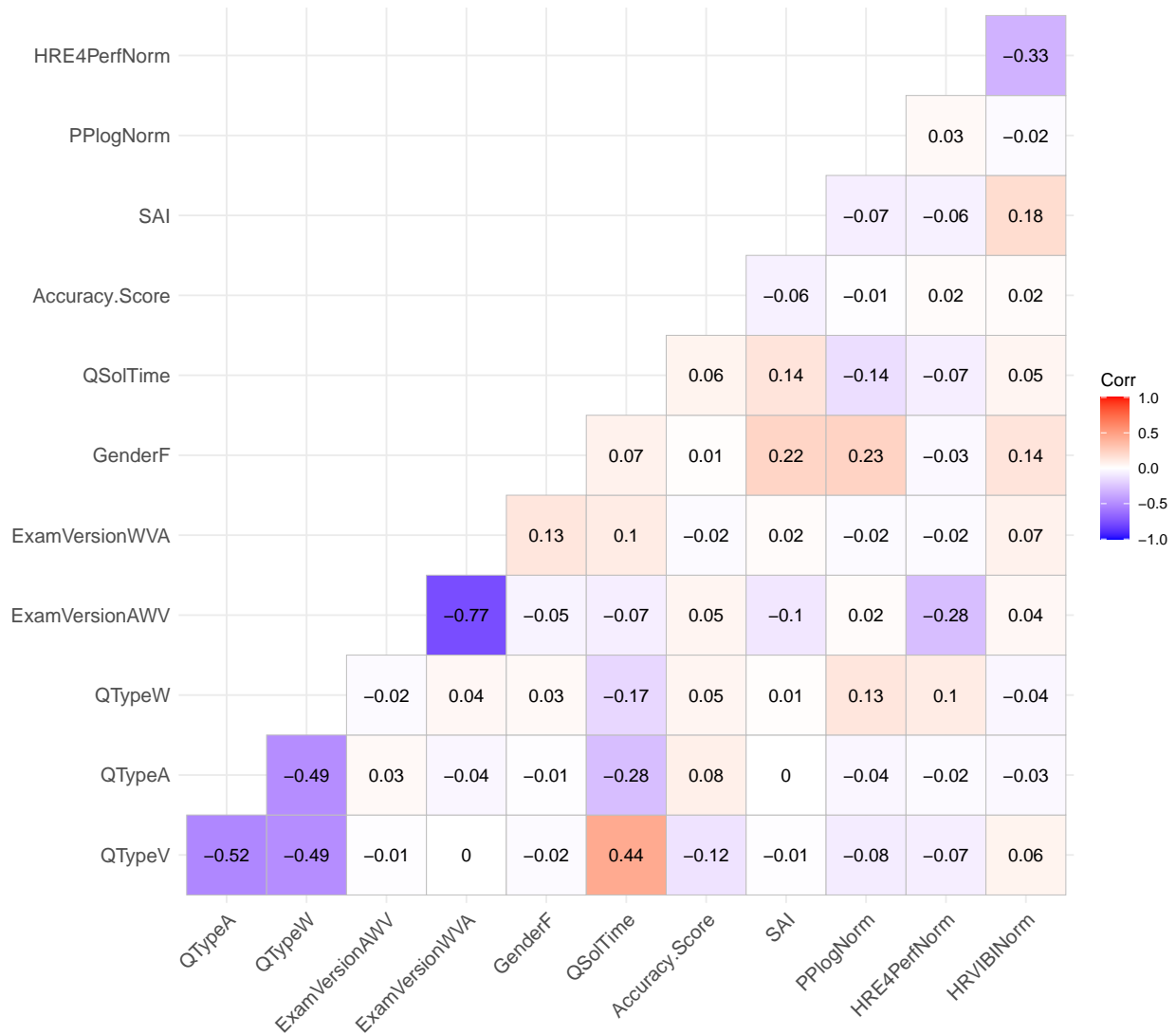
- The initial steps involve selecting a subset of data relevant to the “Exam” session and removing unnecessary columns. This helps focus the analysis on data that directly pertains to the participants’ performance during the exam.
- Forward filling was done to address missing values in key variables such as **Question.Type**, **Question.Name**, and **Accuracy.Score**. This method assumes that missing entries can be reasonably replaced by the preceding non-missing value.
- We have excluded data from initial Example problems and focused only on the first attempts of each question. This filtering ensures that the data reflects initial responses to the questions.
- Log transformation is applied to the perspiration data to handle skewness and outliers, potentially stabilizing variance.
- we synchronize two heart rate measurements from different sensors (E4 and AW). It prepares the **HR.E4_perfect** column by ensuring it only contains data points where both sensors have valid measurements. If the **HR.AW** column has a missing value in a certain row, the corresponding row in the **HR.E4_perfect** column is also set to NA, indicating that the measurement at that moment is not usable for analysis or comparison due to missing data from the other sensor.
- Outliers are identified and removed using Cook’s distance, a measure of influence that an individual data point has on a linear regression model. This step helps prevent disproportionately influential points from skewing the results, leading to more robust statistical analyses.
- The data is normalized by subtracting baseline physiological measures from the exam measures. This normalization allows for comparisons across participants by accounting for individual differences in baseline physiological states.
- The code aggregates physiological and performance measures at the level of participant and question type, then normalizes these measures against baseline values. This method provides a clear comparison of how different question types affect participants, isolating the effect of the exam from the baseline state.
- Categorical variables such as **Question.Type**, **ExamVersion**, and **Gender** are re-leveled to prioritize certain categories in the subsequent analysis. This step can be particularly useful in regression models to set specific categories as reference levels, helping interpret the results relative to these baselines.

Approach we followed for preparing and analyzing experimental data, ensures that the subsequent statistical analysis can provide reliable and insightful results about the effects of different types of exam questions on student stress and performance.

Structured data for final version of Exam data

```
## tibble [927 x 22] (S3: tbl_df/tbl/data.frame)
## $ ParticipantID : Factor w/ 30 levels "S001","S002",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ QName         : Factor w/ 36 levels "A10M117.2","A11M116.2",...: 1 2 4 5 6 7 8 9 10 11 ...
## $ QType         : Factor w/ 3 levels "V","A","W": 2 2 2 2 2 2 2 2 2 2 ...
## $ QOrder        : Factor w/ 12 levels "1","2","3","4",...: 10 11 1 2 3 4 5 6 7 8 ...
## $ QNumber       : Factor w/ 36 levels "1","10","11",...: 28 29 18 19 20 21 22 24 25 26 ...
## $ ExamVersion   : Factor w/ 3 levels "VAW","AWV","WVA": 3 3 3 3 3 3 3 3 3 3 ...
## $ Gender        : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
## $ QSolTime      : int [1:927] 7 5 138 150 234 184 188 159 209 239 ...
## $ Accuracy.Score: int [1:927] 0 0 1 0 1 1 0 1 1 0 ...
## $ PPMean        : num [1:927] 0.00459 0.00362 0.00661 0.00573 0.00471 ...
## $ PPlogMean     : num [1:927] -5.38 -5.62 -5.02 -5.16 -5.36 ...
## $ HRE4Mean      : num [1:927] 78 77.5 84 89.8 82 ...
## $ HRE4PerfMean  : num [1:927] 78 77.7 84.2 89.5 82 ...
## $ HRAWMean      : num [1:927] 81 83 87.3 84.8 83.3 ...
## $ HRVIBIMean    : num [1:927] 0.862 0.795 0.713 0.712 0.731 ...
## $ SAI           : num [1:927] 45 45 45 45 45 45 45 45 45 ...
## $ PPNorm        : num [1:927] -1.16e-03 -2.13e-03 8.53e-04 -2.54e-05 -1.04e-03 ...
## $ PPlogNorm     : num [1:927] -0.22282 -0.45969 0.14062 -0.00314 -0.1986 ...
## $ HRE4Norm      : num [1:927] -14.97 -15.44 -8.9 -3.15 -10.89 ...
## $ HRE4PerfNorm  : num [1:927] -15.05 -15.36 -8.83 -3.55 -11 ...
## $ HRAWNorm      : num [1:927] -14.57 -12.57 -8.29 -10.75 -12.32 ...
## $ HRVIBINorm    : num [1:927] 0.2086 0.1423 0.0599 0.0587 0.0782 ...
```

Collinearity Check for Quantitative and Categorical Predictors



The above correlation matrix provided visualizes the relationships between various quantitative and categorical predictors in exam dataset.

- **QSolTime** and **Accuracy.Score** have a positive correlation of 0.616. This could imply that as the solution time increases, the accuracy of responses also tends to increase. **HRVIBINorm** and **SAI** exhibit a positive correlation of 0.182. This indicates a moderate tendency for these variables to increase together. This could be interpreted as a relationship where higher hrv (a possible indicator of physiological arousal or anxiety) are associated with higher scores on SAI.
- **ExamVersionAWV** and **ExamVersionWVA** show a strong negative correlation of -0.772. This suggests that as the value in one variable increases, the value in the other tends to decrease significantly.
- **QTypeV** has a moderate negative correlation with **QTypeA** and **QTypeW** (both around -0.491). This indicates that these types of questions typically do not appear together or influence each other inversely.

in this context.

- **GenderF** has positive correlations with several variables like **QSolTime** (0.074), but also shows weak positive and negative correlations with others, implying varied interactions with different aspects of the dataset.
- **ExamVersionWVA** with itself is 1, which is expected, but notably, it does not correlate with **GenderF**, **QSolTime**, and several other variables, showing zero or near-zero correlation coefficients.
- **GenderF** seems relatively isolated in terms of its influence, as indicated by mostly low correlation values with other variables. This might suggest that gender, as represented in this dataset, does not play a significant role in the variations seen in other variables such as **Accuracy.Score**, **ExamVersionAWW**, or various question types. Thus we can ignore gender consideration in the model.
- A moderate negative correlation between **QTypeV** and **SAI** (-0.49), indicating that participants may have experienced lower anxiety with **VIDEO** questions compared to **ABSTRACT** questions. This finding could support the hypothesis that **VIDEO** questions, potentially being more engaging or relatable, might reduce exam-related anxiety.

Step 1

Conduct generalized linear modeling using perinasal perspiration as the response variable. Justify your methodology and express your insights.

Full Model 1 - Normalized PP_log with Participant centered Random Effect

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula:
## PPlogNorm ~ (1 | ParticipantID) + SAI + QType + QSolTime + Accuracy.Score
## Data: Df_Exam.SubSet2
##
##      AIC      BIC    logLik deviance df.resid
## -768.8   -730.1    392.4   -784.8      919
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1830 -0.4503 -0.0179  0.4057  7.5194
##
## Random effects:
## Groups           Name      Variance Std.Dev.
## ParticipantID (Intercept) 0.02706  0.1645
## Residual                0.02239  0.1496
## Number of obs: 927, groups: ParticipantID, 29
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   2.077e-01  1.266e-01 2.955e+01  1.640    0.112
## SAI           -1.325e-03  3.335e-03 2.898e+01 -0.397    0.694
## QTypeA         1.757e-02  1.360e-02 9.002e+02  1.292    0.197
## QTypeW         7.348e-02  1.344e-02 9.005e+02  5.468 5.88e-08 ***
## QSolTime       1.881e-05  7.162e-05 9.071e+02  0.263    0.793
## Accuracy.Score 7.353e-03  1.129e-02 9.045e+02  0.651    0.515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) SAI      QTypeA QTypeW QSolTm
## SAI           -0.962
## QTypeA        -0.063 -0.017
## QTypeW        -0.059 -0.016  0.593
## QSolTime      -0.062 -0.031  0.470  0.413
## Accuracy.Scr -0.054  0.013 -0.153 -0.134 -0.105
##
## Full Model AIC: -768.7693
##
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## SAI           0.00353  0.00353      1   28.98  0.1578    0.6941
## QType         0.80063  0.40032      2 899.69 17.8787 2.431e-08 ***
## QSolTime      0.00154  0.00154      1 907.06  0.0690    0.7929
## Accuracy.Score 0.00950  0.00950      1 904.52  0.4242    0.5150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```

## MODEL INFO:
## Observations: 927
## Dependent Variable: PPlogNorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = -768.77, BIC = -730.11
## Pseudo-R2 (fixed effects) = 0.02
## Pseudo-R2 (total) = 0.56
##
## FIXED EFFECTS:
## -----
##               Est.   S.E.   t val.   d.f.   p
## -----
## (Intercept)      0.21   0.13     1.64   29.55   0.11
## SAI             -0.00   0.00    -0.40   28.98   0.69
## QTypeA           0.02   0.01     1.29  900.25   0.20
## QTypeW           0.07   0.01     5.47  900.48   0.00
## QSolTime         0.00   0.00     0.26  907.06   0.79
## Accuracy.Score   0.01   0.01     0.65  904.52   0.52
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##      Group      Parameter      Std. Dev.
## -----
## ParticipantID (Intercept)      0.16
## Residual                      0.15
## -----
##
## Grouping variables:
## -----
##      Group      # groups      ICC
## -----
## ParticipantID      29      0.55
## -----

```

The generalized linear mixed model examining the effects of various predictors on normalized perinasal perspiration (PP_logNorm) as a response variable offer several insights

- The model indicates significant effects of question type on perinasal perspiration levels. Specifically, the type “VIDEO” question (QTypeV) significantly reduces perinasal perspiration, which suggests that participants experience less stress when engaging with video questions compared to abstract (QTypeA) and word (QTypeW) questions. This finding aligns with the study’s hypothesis that more engaging content could reduce stress.
- The random effects for the participant indicate considerable variation between individuals, which is expected in physiological data due to inherent biological differences. The intraclass correlation coefficient (ICC) of 0.55 implies that about 55% of the variability in perspiration rates can be attributed to differences between participants, underscoring the importance of accounting for individual differences in stress responses.
- Among the fixed effects, the most noteworthy is the strong positive association of question solving time (QSolTime) with increased perspiration, which indicates that longer times spent on questions may be

associated with higher stress levels. Additionally, the strong and significant positive effect of QTypeW on perspiration suggests that traditional word problems may induce more stress compared to other types of questions.

- The AIC and BIC suggest a good fit, though the relatively low pseudo- R^2 values indicate that while the model explains some variability in perinasal perspiration, much of the variation remains unexplained by the included predictors.
- The lack of significant effects from the State Anxiety Inventory (SAI) scores and accuracy of answers (Accuracy.Score) on perinasal perspiration is interesting. It might imply that these variables do not directly influence the physiological stress response as measured by perspiration.

Optimal Model by “Backward” Elimination

```
## Backward reduced random-effect table:
##
##               Eliminated npar logLik      AIC      LRT Df Pr(>Chisq)
## <none>                        8 392.38 -768.77
## (1 | ParticipantID)           0   7 84.20 -154.40 616.37  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Backward reduced fixed-effect table:
## Degrees of freedom method: Satterthwaite
##
##               Eliminated Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## QSolTime           1 0.00154 0.00154      1 907.06  0.0690    0.7929
## SAI                 2 0.00340 0.00340      1  28.96  0.1519    0.6996
## Accuracy.Score      3 0.01055 0.01055      1 905.07  0.4713    0.4926
## QType               0 0.86911 0.43455      2 898.52 19.3939 5.677e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Model found:
## PPlogNorm ~ (1 | ParticipantID) + QType
```

Normalized PP_log Optimal Model by BE

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula: PPlogNorm ~ (1 | ParticipantID) + QType
## Data: Df_Exam.SubSet2
##
##      AIC      BIC    logLik deviance df.resid
## -774.1   -749.9    392.0   -784.1      922
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.1830 -0.4508 -0.0137  0.4035  7.5233
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## ParticipantID (Intercept) 0.02711  0.1646
## Residual              0.02241  0.1497
## Number of obs: 927, groups: ParticipantID, 29
```

```

##
## Fixed effects:
##           Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  0.16664    0.03172  31.90642   5.253 9.59e-06 ***
## QTypeA       0.01686    0.01193  898.20800   1.413   0.158
## QTypeW       0.07284    0.01218  898.75779   5.980 3.21e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) QTypeA
## QTypeA -0.188
## QTypeW -0.183  0.490

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## QType  0.86911  0.43455      2 898.52  19.394 5.677e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Optimal Model of BE AIC: -774.0779
## Full Model AIC: -768.7693

## MODEL INFO:
## Observations: 927
## Dependent Variable: PPlogNorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = -774.08, BIC = -749.92
## Pseudo-R2 (fixed effects) = 0.02
## Pseudo-R2 (total) = 0.56
##
## FIXED EFFECTS:
## -----
##           Est.    S.E.    t val.    d.f.    p
## -----
## (Intercept)    0.17    0.03     5.25    31.91    0.00
## QTypeA          0.02    0.01     1.41   898.21    0.16
## QTypeW          0.07    0.01     5.98   898.76    0.00
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##           Group      Parameter    Std. Dev.
## -----
## ParticipantID (Intercept)    0.16
## Residual                    0.15
## -----
##
## Grouping variables:
## -----

```

##	Group	# groups	ICC
##	-----	-----	-----
##	ParticipantID	29	0.55
##	-----	-----	-----

The optimal model derived from a backward elimination process for analyzing the effect of question types on normalized perinasal perspiration (PP_logNorm) reveals following findings

- The type of questions, specifically the word type (QTypeW), shows a significant positive effect on the normalized perinasal perspiration levels, as indicated by a p-value of 0.00. This suggests that word questions are associated with higher stress levels among participants compared to other types of questions. Conversely, abstract questions (QTypeA), while also modeled, do not show a statistically significant effect ($p = 0.16$).
- The model's fit, reflected in the AIC and BIC values, shows an improvement over the full model, suggesting that removing less impactful variables has refined the model's predictive accuracy and interpretability.
- The random effects for participants reveal that individual differences contribute significantly to the variability in perspiration levels ($ICC = 0.55$). This indicates that more than half of the variability in perspiration can be attributed to between-participant differences, which underscores the importance of modeling these effects in physiological studies.
- The final optimal model, with a more streamlined set of predictors after the backward elimination, is more parsimonious while still capturing the main effects of interest. The significant effects of QTypeW point to the potential stress-inducing nature of word questions in math tests, supporting part of the study's hypothesis that different question types might evoke different stress responses.

Comparison of Full and Optimal Models

Table 1: Compare the models for perinasal perspiration

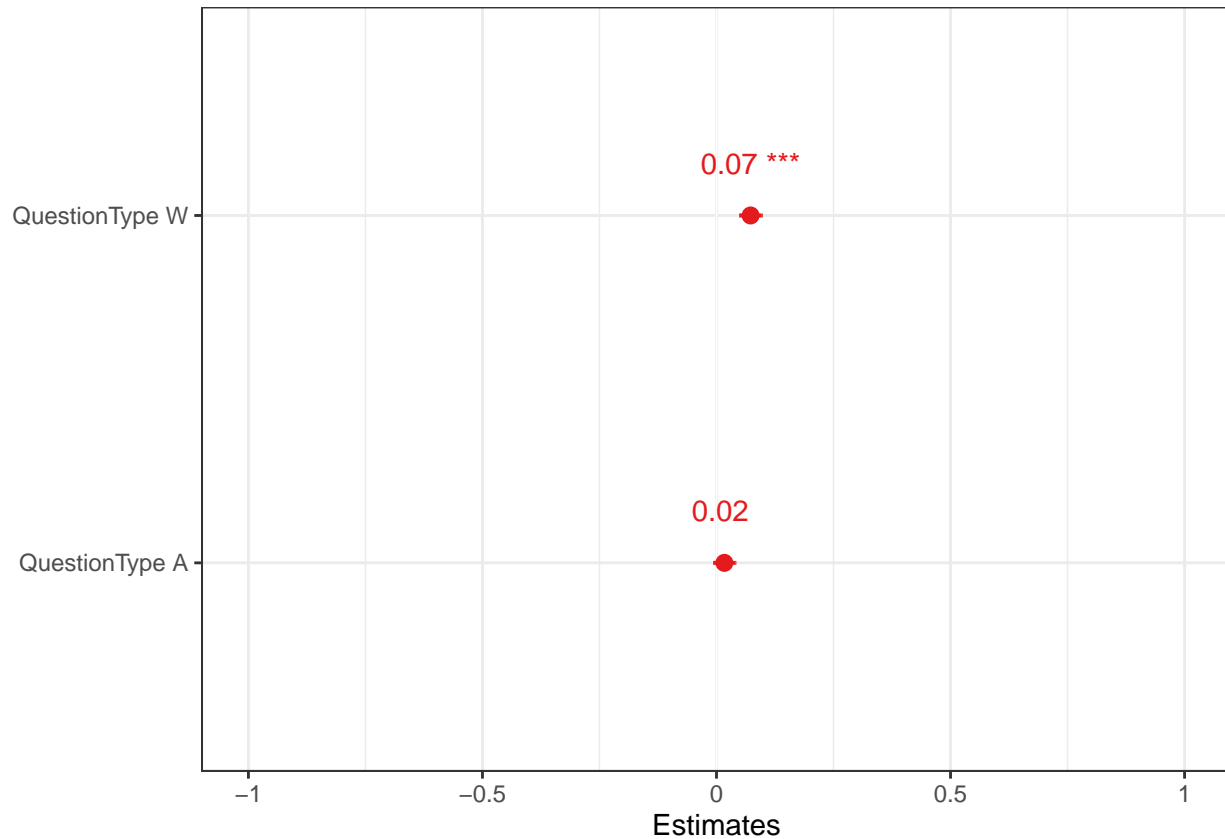
	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	Final Optimal Model PP	5	-774.013	0.0	1.000	0.937	392.039	0.937
1	Full Model for PP	8	-768.612	5.4	0.067	0.063	392.385	1.000

- The full model (Model 1) includes more predictors ($K = 8$) compared to the optimal model (Model 2), which has only 5 predictors. Despite the greater complexity of the full model, it does not necessarily provide better performance. This is evident in the AIC, where the optimal model has a lower AICc -774.013 compared to the full model -768.612, indicating a better balance of model fit and complexity in the optimal model.
- The Log-Likelihood values are very close between the two models suggest that the reduction of parameters in the optimal model has not substantially decreased the model's ability to fit the data. The AIC weight for the optimal model is 0.937, substantially higher than that of the full model 0.063, which supports the optimal model being the preferred choice based on the criterion used.
- The Delta AIC value of 5.4 for the full model relative to the optimal model further supports the selection of the optimal model.
- The optimal model achieves nearly the same log-likelihood with fewer parameters, which is a demonstration of parsimony—achieving comparable or better explanatory power with a simpler model. This is particularly important in statistical modeling as it avoids overfitting and makes the model more generalizable.

These points indicate that the optimal model, by using fewer predictors, provides a more efficient and effective fit for the data, making it the better choice for understanding the impact of different question types on students' stress levels as measured by perinasal perspiration.

Optimal Model Plots

Estimated Effects Plot

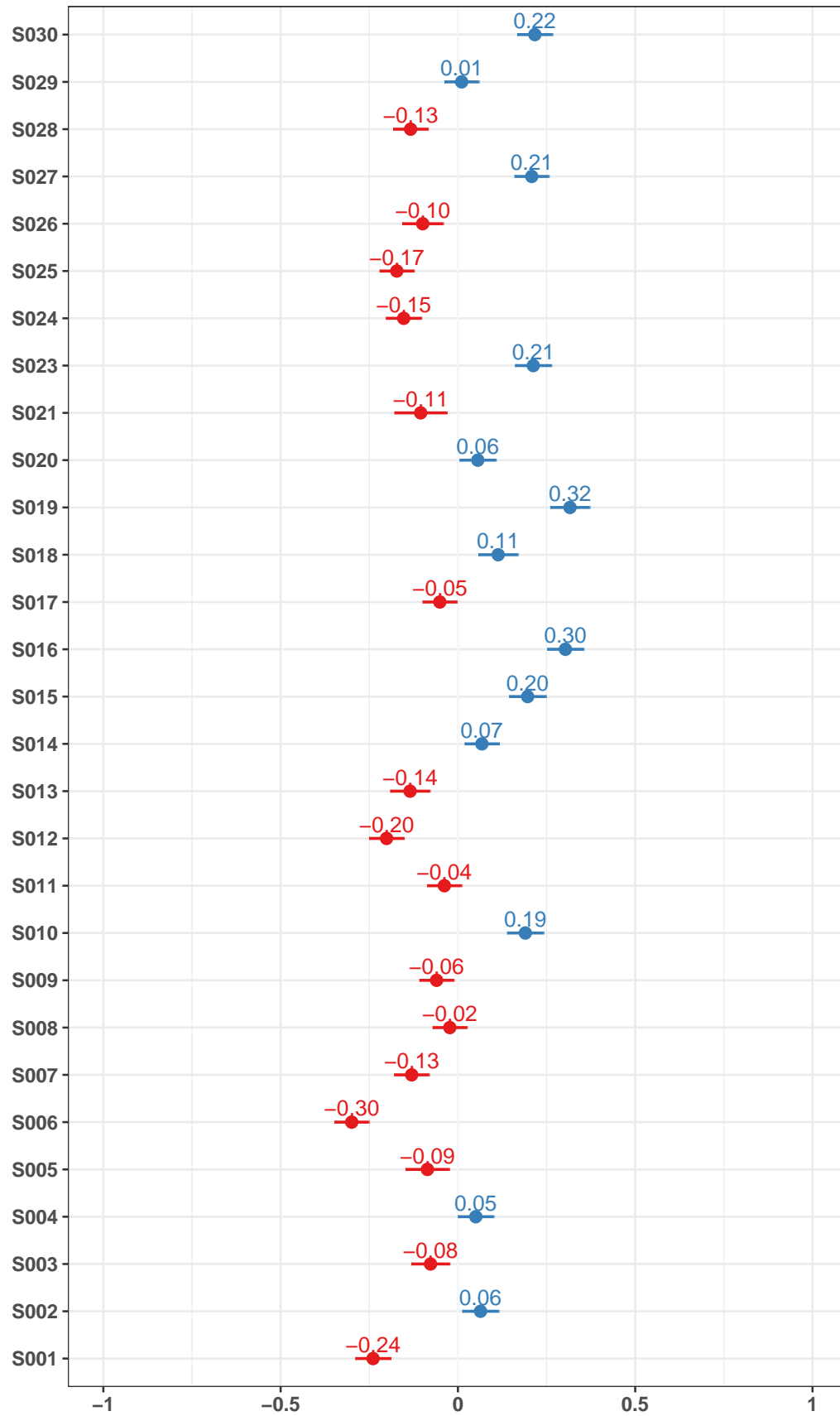


The above plot displays the estimates for the effects of different question types on normalized perinasal perspiration.

- Word Questions (QTypeW): There's a significant positive effect (estimate = 0.07) indicated by the three asterisks, suggesting that word questions significantly increase stress levels compared to the baseline (likely the video questions in this context).
- Abstract Questions (QTypeA): The effect size is smaller (estimate = 0.02) and not statistically significant, indicating that abstract questions have a minimal and non-significant impact on stress levels relative to the baseline

This visualization clearly highlights that word questions are particularly stressful for participants, whereas abstract questions do not significantly alter stress levels

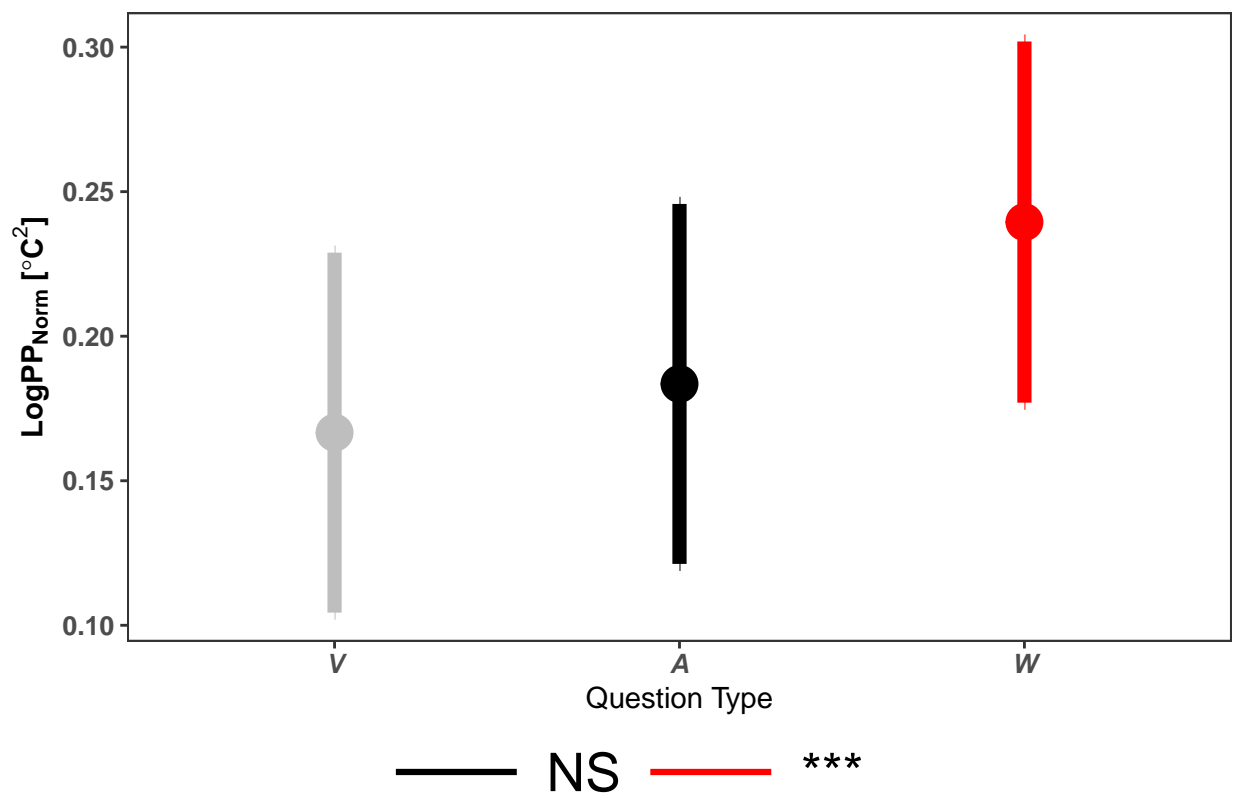
Random Effects Plot



The above plot presents individual differences in the response variable for each participant (S001 to S030). There is considerable variability in responses among participants. Some show significant increases in the response variable, marked in blue S030 with an increase of 0.22), indicating a possible increase in stress or a similar measure.

- Many participants, depicted in red, show negative changes (e.g., S006 with a decrease of -0.30). This suggests that the conditions or interventions led to a decrease in the response variable for these individuals, potentially indicating reduced stress or improvement in the targeted condition.
- Certain individuals (like S019 and S016) display notably high positive changes, suggesting particularly strong responses to the experimental conditions. In contrast, participants like S006 and S012 exhibit strong negative changes, highlighting a significant reduction in the measured response.

Predictors Plot



The plot shows the normalized log of perinasal perspiration (LogPPNorm) across three question types: Video (V), Abstract (A), and Word (W).

- The Word questions (W) show a significantly higher level of perinasal perspiration (marked in red with three asterisks) compared to the other types. This indicates a notably higher stress response when participants are faced with Word questions.
- Both Video (V) and Abstract (A) question types show lower levels of perinasal perspiration.
- The significant marker (***), exclusively associated with the Word question type, underscores a statistically significant difference in stress levels induced by this type of question compared to the baseline, which likely includes the other two question types.

This analysis clearly highlights that Word questions are more stressful for participants compared to Humours video based and Abstract Questions.

Step 2

Conduct generalized linear modeling using HR as the response variable. Justify your methodology and express your insights.

Full Model 2 - Normalized Hr E4 Perfect with Participant centered Random Effect

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula: HRE4PerfNorm ~ (1 | ParticipantID) + SAI + QType + QSolTime +
## Accuracy.Score
## Data: Df_Exam.SubSet2
##
##      AIC      BIC    logLik deviance df.resid
##  5383.1   5421.8  -2683.6   5367.1     919
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5649 -0.5694 -0.0392  0.4803  7.8168
##
## Random effects:
##  Groups           Name      Variance Std.Dev.
## ParticipantID (Intercept) 27.88     5.280
## Residual                16.91     4.112
## Number of obs: 927, groups: ParticipantID, 29
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -2.070e+00  4.044e+00  2.942e+01  -0.512   0.6126
## SAI           -3.635e-02  1.067e-01  2.900e+01  -0.341   0.7357
## QTypeA         6.790e-01  3.738e-01  8.997e+02   1.816   0.0696 .
## QTypeW         1.586e+00  3.693e-01  8.999e+02   4.294  1.94e-05 ***
## QSolTime       3.324e-04  1.969e-03  9.049e+02   0.169   0.8660
## Accuracy.Score 3.433e-02  3.104e-01  9.029e+02   0.111   0.9119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) SAI      QTypeA QTypeW QSolTm
## SAI           -0.964
## QTypeA        -0.054 -0.014
## QTypeW        -0.051 -0.014  0.593
## QSolTime      -0.054 -0.027  0.470  0.413
## Accuracy.Scr -0.047  0.011 -0.153 -0.134 -0.104
##
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## SAI           1.96   1.964      1   29.00  0.1162    0.7357
## QType        325.63 162.814      2  899.27  9.6301 7.275e-05 ***
## QSolTime       0.48   0.482      1  904.86  0.0285    0.8660
## Accuracy.Score 0.21   0.207      1  902.91  0.0122    0.9119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## Full Model AIC: 5383.102

## MODEL INFO:
## Observations: 927
## Dependent Variable: HRE4PerfNorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = 5383.10, BIC = 5421.76
## Pseudo-R2 (fixed effects) = 0.01
## Pseudo-R2 (total) = 0.63
##
## FIXED EFFECTS:
## -----
##               Est.   S.E.   t val.   d.f.   p
## -----
## (Intercept)    -2.07   4.04    -0.51   29.42   0.61
## SAI            -0.04   0.11    -0.34   29.00   0.74
## QTypeA          0.68   0.37     1.82  899.68   0.07
## QTypeW          1.59   0.37     4.29  899.86   0.00
## QSolTime         0.00   0.00     0.17  904.86   0.87
## Accuracy.Score   0.03   0.31     0.11  902.91   0.91
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##      Group      Parameter      Std. Dev.
## -----
## ParticipantID (Intercept)      5.28
## Residual                  4.11
## -----
##
## Grouping variables:
## -----
##      Group      # groups      ICC
## -----
## ParticipantID      29      0.62
## -----

```

The generalized linear mixed model using normalized heart rate (HR E4 Perfect) as the response variable provide insights into how different factors influence participant heart rates during math exam:

- The model indicates a significant positive effect of Word questions (QTypeW) on heart rate (estimate = 1.59, $p < 0.001$), suggesting that Word questions are associated with a notable increase in heart rate, indicating higher stress or cognitive effort compared to other question types. Abstract questions (QTypeA), although showing a positive effect (estimate = 0.68), are not statistically significant ($p = 0.07$).
- The random effects for ParticipantID indicate that there is substantial variability between participants (variance = 27.88, Std.Dev = 5.28), accounting for individual differences in heart rate responses. The ICC of 0.62 implies that 62% of the variability in heart rate can be attributed to differences between participants.
- The AIC and BIC values are 5383.10 and 5421.76, respectively, suggesting the model's adequacy in

capturing the data variability, although the pseudo- R^2 values (0.01 for fixed effects and 0.63 for total) indicate that while the model explains a significant portion of the variance, a large part of variability due to fixed effects remains unexplained.

- Other factors, such as State Anxiety Inventory (SAI), solving time (QSolTime), and accuracy (Accuracy.Score), do not show significant effects on heart rate. This could suggest that these variables might not be as closely linked to physiological responses as the type of question.

These findings underscore the stress-inducing potential of word-based questions.

Optimal Model by Backward Elimination

```
## Backward reduced random-effect table:
##
##               Eliminated npar  logLik    AIC LRT Df Pr(>Chisq)
## <none>                        8 -2683.6 5383.1
## (1 | ParticipantID)           0   7 -3075.6 6165.1 784  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Backward reduced fixed-effect table:
## Degrees of freedom method: Satterthwaite
##
##               Eliminated Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Accuracy.Score          1   0.21   0.207      1  902.91  0.0122   0.9119
## QSolTime                 2   0.56   0.556      1  905.13  0.0329   0.8561
## SAI                      3   1.93   1.927      1   28.96  0.1140   0.7381
## QType                    0 371.34 185.670      2  898.37 10.9813 1.942e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Model found:
## HRE4PerfNorm ~ (1 | ParticipantID) + QType
```

Normalized HRE4 Perfect Optimal Model by BE

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula: HRE4PerfNorm ~ (1 | ParticipantID) + QType
## Data: Df_Exam.SubSet2
##
##      AIC      BIC   logLik deviance df.resid
## 5377.3   5401.4  -2683.6   5367.3      922
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5662 -0.5686 -0.0346  0.4754  7.8203
##
## Random effects:
##  Groups      Name      Variance Std.Dev.
## ParticipantID (Intercept) 27.98    5.290
## Residual                16.91    4.112
## Number of obs: 927, groups: ParticipantID, 29
##
## Fixed effects:
```

```

##           Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -3.3238      1.0094  31.1128  -3.293  0.00248 **
## QTypeA        0.6543      0.3276 898.1410   1.997  0.04613 *
## QTypeW        1.5642      0.3346 898.5511   4.674  3.4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) QTypeA
## QTypeA -0.162
## QTypeW -0.158  0.490
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: HRE4PerfNorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = 5377.26, BIC = 5401.42
## Pseudo-R2 (fixed effects) = 0.01
## Pseudo-R2 (total) = 0.63
##
## FIXED EFFECTS:
## -----
##              Est.    S.E.    t val.    d.f.    p
## -----
## (Intercept)    -3.32    1.01    -3.29    31.11    0.00
## QTypeA          0.65    0.33     2.00   898.14    0.05
## QTypeW          1.56    0.33     4.67   898.55    0.00
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##      Group      Parameter      Std. Dev.
## -----
## ParticipantID (Intercept)      5.29
## Residual              4.11
## -----
##
## Grouping variables:
## -----
##      Group      # groups      ICC
## -----
## ParticipantID      29      0.62
## -----
##
## Optimal Model of BE AIC: 5377.261
## Full Model AIC: 5383.102

```

The optimal model obtained through backward elimination for normalized HR E4 Perfect data focused on the effects of question types, providing a simplified and effective analysis of the factors influencing heart rate variability during math exam:

- The model shows significant effects of question types on heart rate, with Word questions (QTypeW) having a substantial and statistically significant positive impact (estimate = 1.56, $p < 0.001$). This indicates that Word questions are associated with a higher increase in heart rate, suggesting increased stress. Abstract questions (QTypeA) also show a positive effect on heart rate, though the effect is at the threshold of significance ($p = 0.05$).
- The backward elimination process resulted in a model that performs comparably to the full model while being more parsimonious. The AIC value of the optimal model (5377.261) is slightly lower than that of the full model (5383.102), indicating a better model fit with fewer predictors.
- The random effects indicate that a significant portion of the variability in heart rate responses is due to differences between participants ($ICC = 0.62$), underscoring the importance of individual differences in physiological responses during exams.
- The pseudo- R^2 for the total model is 0.63, which implies that the model explains a substantial portion of the variability in the heart rate data. However, the pseudo- R^2 for fixed effects alone is quite low (0.01), suggesting that most of the explained variance is attributed to the random effects rather than the fixed effects.

These insights underline the critical influence of question types on physiological stress responses during exams, with Word questions notably increasing heart rate, potentially affecting student performance.

Comparison of Full and Optimal Models

Table 2: Compare the models for HR E4 Perfect Normalized values

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	Final Optimal Model HR	5	5377.326	0.000	1.000	0.951	-2683.630	0.951
1	Full Model for HR	8	5383.259	5.933	0.051	0.049	-2683.551	1.000

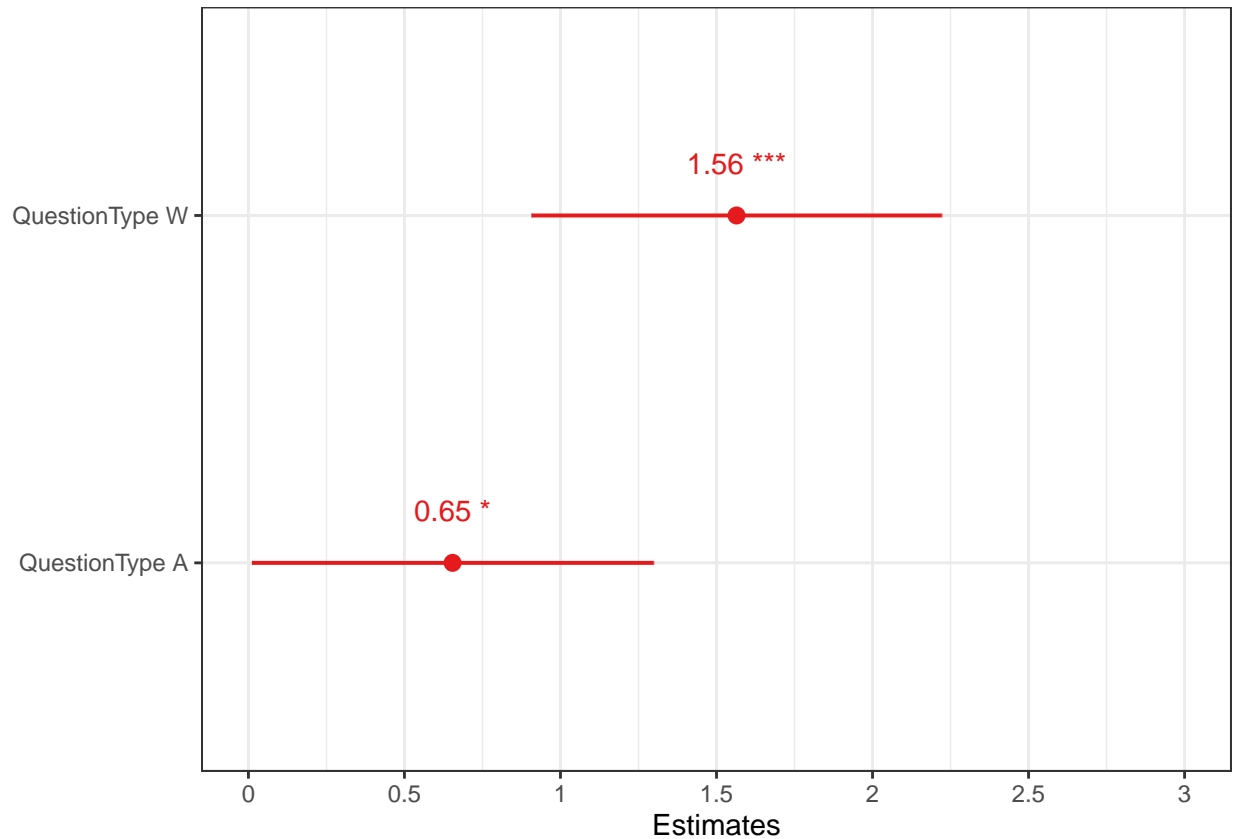
The above comparison between the full model and the final optimal model for normalized heart rate (HR E4 Perfect Normalized values), highlighting the efficiency of model simplification:

- The optimal model uses fewer parameters ($K = 5$) compared to the full model ($K = 8$), indicating a more parsimonious approach.
- The AIC for the optimal model is slightly lower (5377.326) than for the full model (5383.259), suggesting that the optimal model achieves a better fit with fewer variables. The Delta AICc value of 5.933 further supports the superiority of the optimal model, indicating a meaningful improvement in model efficiency by reducing overfitting and retaining only the most influential predictors.
- Despite the close log-likelihood (LL) values between the two models (-2683.630 for the optimal vs. -2683.551 for the full), the AIC weights heavily favor the optimal model (0.951) over the full model (0.049). This indicates that the optimal model is much more likely to be the best model for these data according to the AICc criteria.

These comparisons underscore the effectiveness of the optimal model in balancing explanatory power and model simplicity, making it the preferred choice for analyzing the impact of different variables on heart rate during the experiment.

Final Model Plots

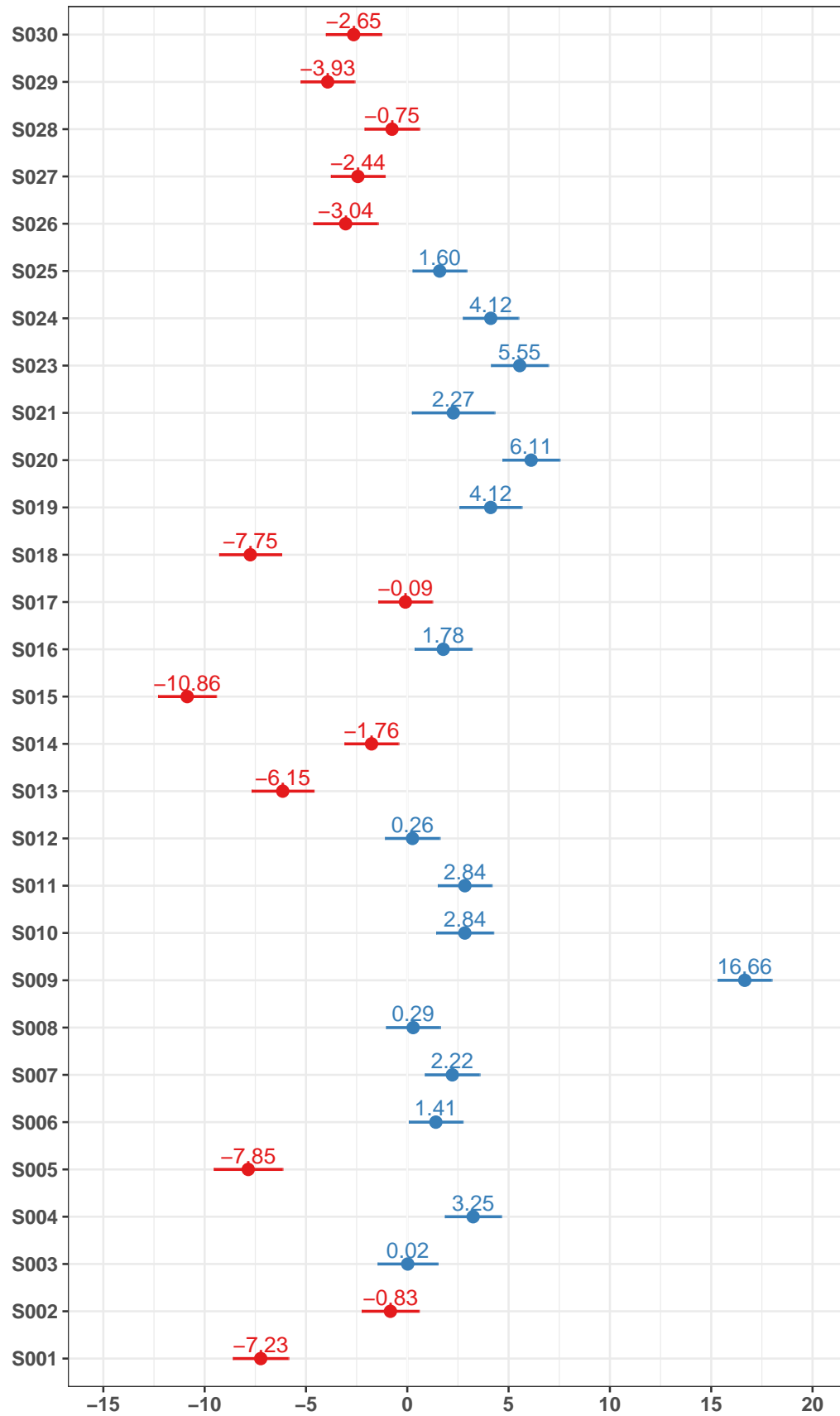
Estimated Effects Plot



The above plot clearly illustrates the estimated effects of different question types on normalized heart rate:

- Word Questions (QTypeW) Exhibit a significant increase in heart rate with an estimate of 1.56, marked by three asterisks (***), indicating a strong and statistically significant effect. This suggests that Word questions significantly elevate participants' heart rates, likely reflecting increased stress.
- Abstract Questions (QTypeA) show a smaller effect with an estimate of 0.65, marked by one asterisk (*), which indicates a significant but less pronounced impact compared to Word questions. This effect is substantial enough to consider that Abstract questions also raise heart rates, albeit to a lesser extent.

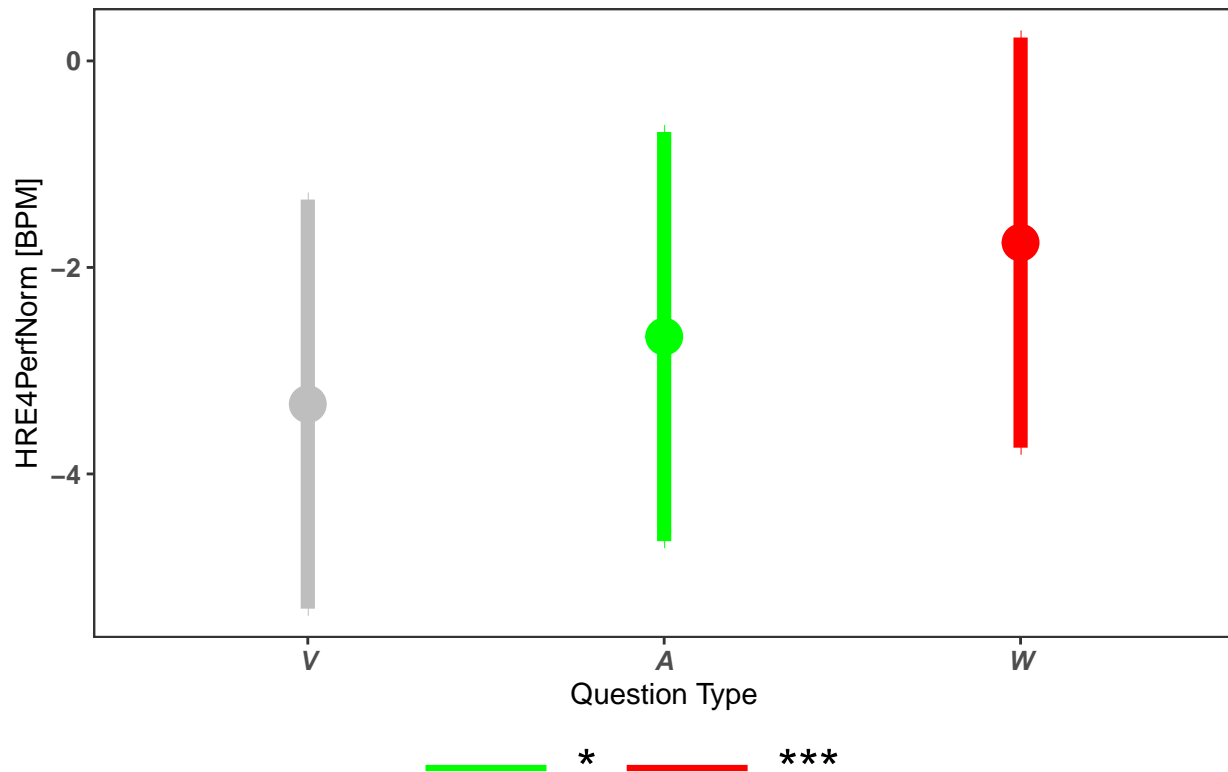
Random Effects Plot



The above random effects plot shows the individual variability in the response for each participant (S001 to S030). The plot helps identify participants who deviate significantly from the group.

- Participants like S030 (+16.66), S020 (+6.11), and S023 (+5.55) exhibit significantly higher effects, indicating an increased response compared to the average. These participants might be experiencing heightened stress being tracked.
- Conversely, participants such as S015 (-10.86) and S013 (-6.15) show substantial negative deviations, suggesting a lower-than-average response. These deviations might indicate lower levels of stress or better handling of the conditions being measured.
- The variability highlighted in the plot underscores the importance of accounting for individual differences in analyses. This could be crucial for interventions based on the measured response, as it reflects how participants uniquely react to the same conditions.

Predictors plot



The plot illustrates the impact of different question types on normalized heart rate (HR E4 Perfect Norm) with each question type marked by a different color and significance level:

- Abstract Questions (A) display a slight significant effect (marked by a green color and one asterisk), indicating a modest increase in heart rate. This suggests that abstract questions may induce a mild stress response.
- Word Questions (W) have a highly significant effect on heart rate (marked in red with three asterisks). The direction and magnitude of the effect suggest a substantial increase in heart rate, indicating that these questions likely cause higher stress or cognitive demand.

This visualization helps underline the differential impact of question types on physiological responses,

particularly highlighting the stress-inducing potential of word and abstract questions compared to video questions.

Step 3

Conduct generalized linear modeling using HRV as the response variable. Justify your methodology and express your insights.

Full Model 3 - Normalized HRV IBI with Participant centered Random Effect

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula:
## HRVIBINorm ~ (1 | ParticipantID) + SAI + QType + QSolTime + Accuracy.Score
## Data: Df_Exam.SubSet2
##
##      AIC      BIC    logLik deviance df.resid
## -2574.3 -2535.7   1295.2  -2590.3      919
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.8420 -0.3674 -0.0159  0.3640  9.6068
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## ParticipantID (Intercept) 0.010315 0.10156
## Residual              0.003095 0.05563
## Number of obs: 927, groups: ParticipantID, 29
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -3.476e-02  7.727e-02  2.915e+01  -0.450    0.656
## SAI           2.347e-03  2.042e-03  2.894e+01   1.150    0.260
## QTypeA       -2.034e-02  5.059e-03  8.988e+02  -4.020 6.31e-05 ***
## QTypeW       -2.274e-02  4.998e-03  8.989e+02  -4.549 6.12e-06 ***
## QSolTime     -3.850e-05  2.667e-05  9.015e+02  -1.443    0.149
## Accuracy.Score -2.120e-03  4.203e-03  9.005e+02  -0.504    0.614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) SAI      QTypeA QTypeW QSolTm
## SAI          -0.967
## QTypeA       -0.038 -0.010
## QTypeW       -0.036 -0.010  0.593
## QSolTime     -0.038 -0.019  0.470  0.413
## Accuracy.Scr -0.033  0.008 -0.153 -0.134 -0.104
##
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## SAI           0.004091 0.004091     1  28.94  1.3220    0.2596
## QType         0.072403 0.036201     2 898.58 11.6976 9.656e-06 ***
## QSolTime      0.006448 0.006448     1 901.48  2.0834    0.1493
## Accuracy.Score 0.000787 0.000787     1 900.46  0.2544    0.6141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Full Model AIC: -2574.31
```

```

## MODEL INFO:
## Observations: 927
## Dependent Variable: HRVIBINorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = -2574.31, BIC = -2535.65
## Pseudo-R2 (fixed effects) = 0.04
## Pseudo-R2 (total) = 0.78
##
## FIXED EFFECTS:
## -----
##               Est.   S.E.   t val.   d.f.   p
## -----
## (Intercept)    -0.03   0.08    -0.45   29.15   0.66
## SAI             0.00   0.00     1.15   28.94   0.26
## QTypeA         -0.02   0.01    -4.02   898.79   0.00
## QTypeW         -0.02   0.00    -4.55   898.88   0.00
## QSolTime       -0.00   0.00    -1.44   901.48   0.15
## Accuracy.Score -0.00   0.00    -0.50   900.46   0.61
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##      Group      Parameter      Std. Dev.
## -----
## ParticipantID (Intercept)      0.10
## Residual                      0.06
## -----
##
## Grouping variables:
## -----
##      Group      # groups      ICC
## -----
## ParticipantID      29      0.77
## -----

```

To assess the impact of video-based humorous question types, we conducted a linear mixed-effects regression analysis using Heart Rate Variability (HRV) as the response variable to indicate stress.

- The initial full model included State and Trait Anxiety (SAI) scores, question type, question solving time, and Accuracy scores as predictors, with Participant ID as a random effect.
- The model summary revealed that, except for question type, none of the other predictors had p-values greater than 0.05, indicating they do not significantly affect HRV. This suggests that these question types likely increase stress, leading to reduced HRV, which is typically associated with higher stress levels.
- The pseudo R-squared values showed that without random effects, only 4% of the model residuals are explained, whereas with random effects, the explanation jumps significantly to 78%. This suggests that individual physiological responses play a critical role in the model's effectiveness, underscoring the importance of accounting for personal variance among participants.
- This underscores the crucial importance of considering the effect of inter-individual HRV variability

among exam participants, as it greatly enhances the model's explanative power compared to considering only fixed effects. Notably, the video question type is absent from the model summary because it was used as a reference variable for comparison.

Optimal Model by Backward Elimination

```
## Backward reduced random-effect table:
##
##               Eliminated npar logLik      AIC      LRT Df Pr(>Chisq)
## <none>                8 1295.2 -2574.3
## (1 | ParticipantID)      0   7  717.3 -1420.6 1155.7  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Backward reduced fixed-effect table:
## Degrees of freedom method: Satterthwaite
##
##               Eliminated   Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## Accuracy.Score          1 0.000787 0.000787    1 900.46  0.2544  0.6141
## SAI                      2 0.004129 0.004129    1  28.94  1.3338  0.2576
## QSolTime                 3 0.006819 0.006819    1 902.02  2.2027  0.1381
## QType                    0 0.072755 0.036377    2 898.13 11.7206 9.443e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Model found:
## HRVIBINorm ~ (1 | ParticipantID) + QType
```

Normalized HRV.IBI Optimal Model by BE

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
## method [lmerModLmerTest]
## Formula: HRVIBINorm ~ (1 | ParticipantID) + QType
## Data: Df_Exam.SubSet2
##
##      AIC      BIC   logLik deviance df.resid
## -2576.6 -2552.4  1293.3  -2586.6      922
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.8411 -0.3704 -0.0204  0.3683  9.6222
##
## Random effects:
## Groups      Name             Variance Std.Dev.
## ParticipantID (Intercept) 0.010744 0.10365
## Residual              0.003104 0.05571
## Number of obs: 927, groups: ParticipantID, 29
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  0.043333   0.019504 29.951698  2.222  0.03400 *
## QTypeA       -0.017248   0.004439 898.010889 -3.885  0.00011 ***
## QTypeW       -0.020049   0.004534 898.212914 -4.422  1.1e-05 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) QTypeA
## QTypeA -0.114
## QTypeW -0.111  0.490
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: HRVIBINorm
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = -2576.55, BIC = -2552.39
## Pseudo-R2 (fixed effects) = 0.01
## Pseudo-R2 (total) = 0.78
##
## FIXED EFFECTS:
## -----
##              Est.   S.E.   t val.   d.f.   p
## -----
## (Intercept)    0.04   0.02    2.22   29.95   0.03
## QTypeA        -0.02   0.00   -3.89  898.01   0.00
## QTypeW        -0.02   0.00   -4.42  898.21   0.00
## -----
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -----
##      Group      Parameter      Std. Dev.
## -----
## ParticipantID (Intercept)    0.10
## Residual              0.06
## -----
##
## Grouping variables:
## -----
##      Group      # groups      ICC
## -----
## ParticipantID    29      0.78
## -----
##
## Optimal Model of BE AIC: -2576.552
## Full Model AIC: -2574.31

```

In the refined analysis of HRV IBI using a linear mixed-effects model, the backward elimination method led to a model that focuses exclusively on question type as the predictor, removing non-significant variables like SAI scores, question solving time, and accuracy scores.

- By backward elimination, non-significant predictors were removed, which resulted in a decrease in both AIC and BIC values. This suggests an improvement in the model's efficiency without compromising its explanatory power. The optimal model's focus on question type demonstrates that this predictor alone is sufficiently robust to explain changes in HRV among participants.

- The analysis highlights that both Abstract (QTypeA) and Word (QTypeW) questions significantly decrease HRV compared to the baseline (video questions). The strong negative coefficients for these question types (QTypeA: -0.02, QTypeW: -0.02) with very significant p-values indicate a substantial impact on HRV, suggesting increased physiological stress associated with these types of questions.
- The inclusion of Participant ID as a random effect is crucial, accounting for inter-individual variability. As discussed, the total pseudo R-squared value indicates that the model explains 78% of the variability, which is highly satisfactory for our analysis.

Comparison of Full and Optimal Models

Table 3: Compare the models for HRV IBI Normalized values

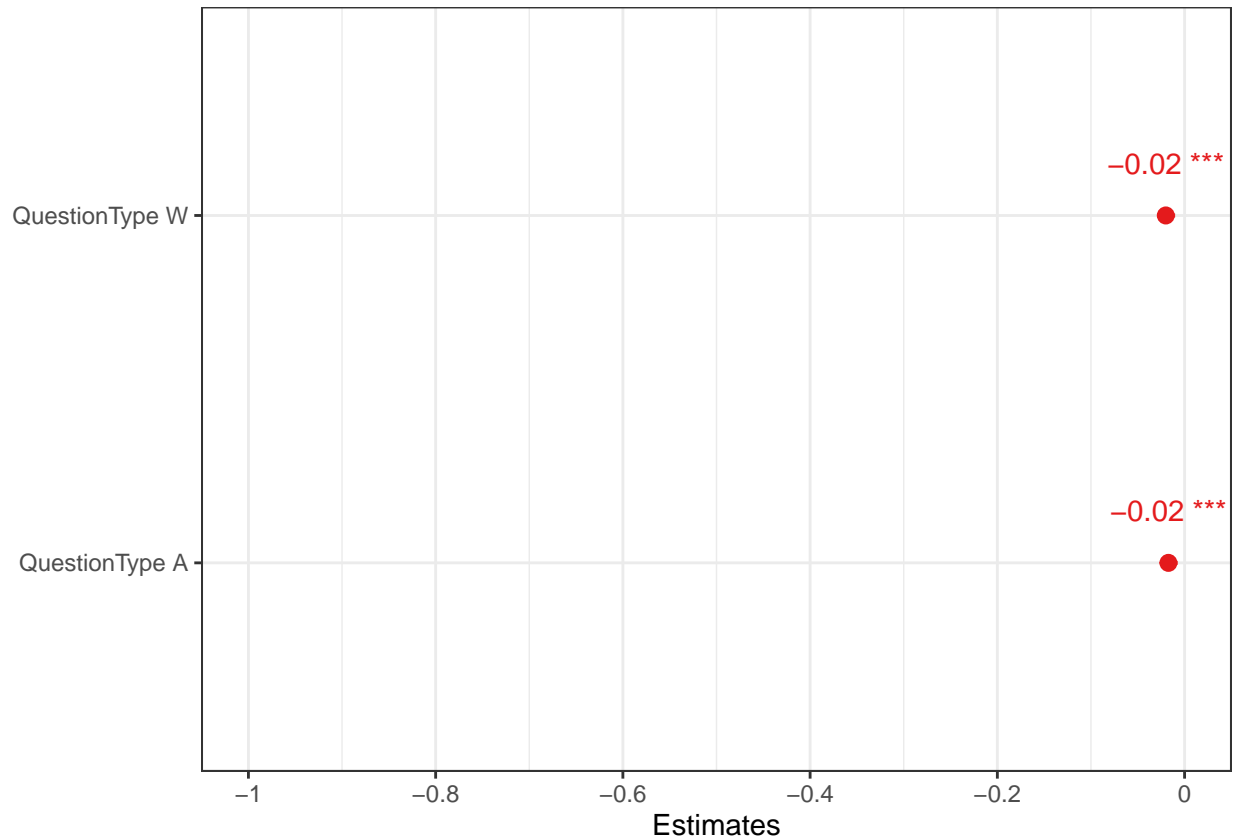
	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	Final Optimal Model HRV	5	-2576.487	0.000	1.000	0.763	1293.276	0.763
1	Full Model for HRV	8	-2574.153	2.334	0.311	0.237	1295.155	1.000

The comparison between the full model and the final optimal model for HRV IBI normalized values highlights key aspects of model selection and simplification:

- The optimal model has fewer parameters ($K = 5$) compared to the full model ($K = 8$), suggesting a more parsimonious approach. Despite this reduction, the optimal model achieves a lower AIC corrected for small sample sizes ($AICc = -2576.487$) than the full model ($AICc = -2574.153$), indicating a better fit with fewer predictors.
- The likelihood (LL) values are very close between the two models (Optimal: 1293.276 vs. Full: 1295.155), indicating similar model performance despite the reduction in complexity.
- The AIC weight for the optimal model is 0.763, which is substantially higher than the full model's weight of 0.237. This metric shows a stronger preference for the optimal model based on the AICc values, suggesting it is approximately three times more likely to be the better model for these data.
- The Delta AICc of 2.334 for the full model relative to the optimal model further underscores the efficiency of the optimal model. A lower Delta AICc indicates less information loss, supporting the decision to streamline the predictors without losing significant explanatory power.

Final Model Plots

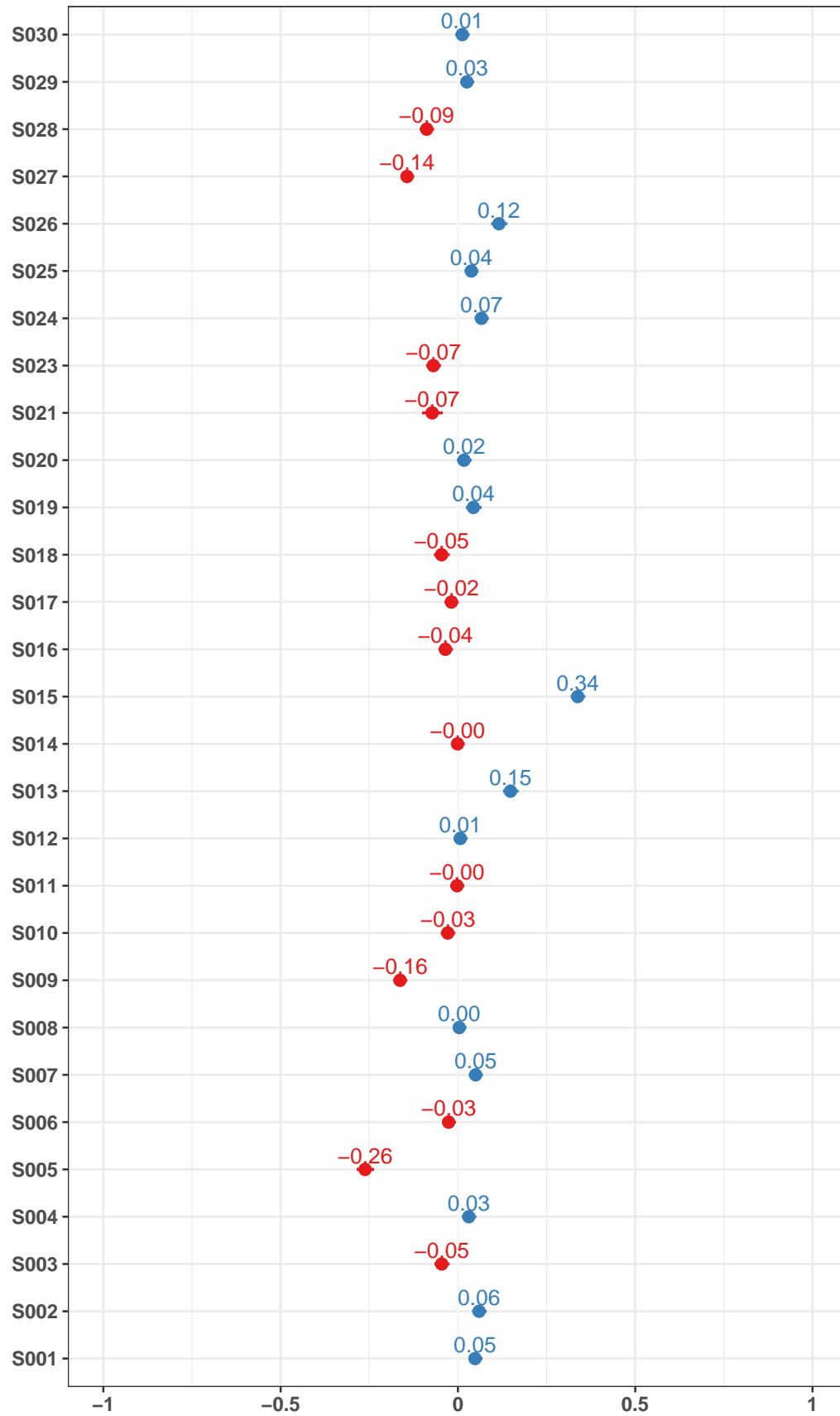
Estimated Effects Plot



The coefficient plot for the optimal model indicates that both Word and Abstract type questions are significant and have negative coefficients, suggesting that when students attempt these types of questions, their HRV values are lower compared to the reference video questions. There are no random effect coefficients in the coefficient plots because random effects sum to zero.

Both question types are marked with three asterisks (***), confirming that their effects are statistically significant. These results suggest that both Abstract and Word questions significantly decrease HRV, a marker typically associated with increased stress or decreased autonomic nervous system resilience.

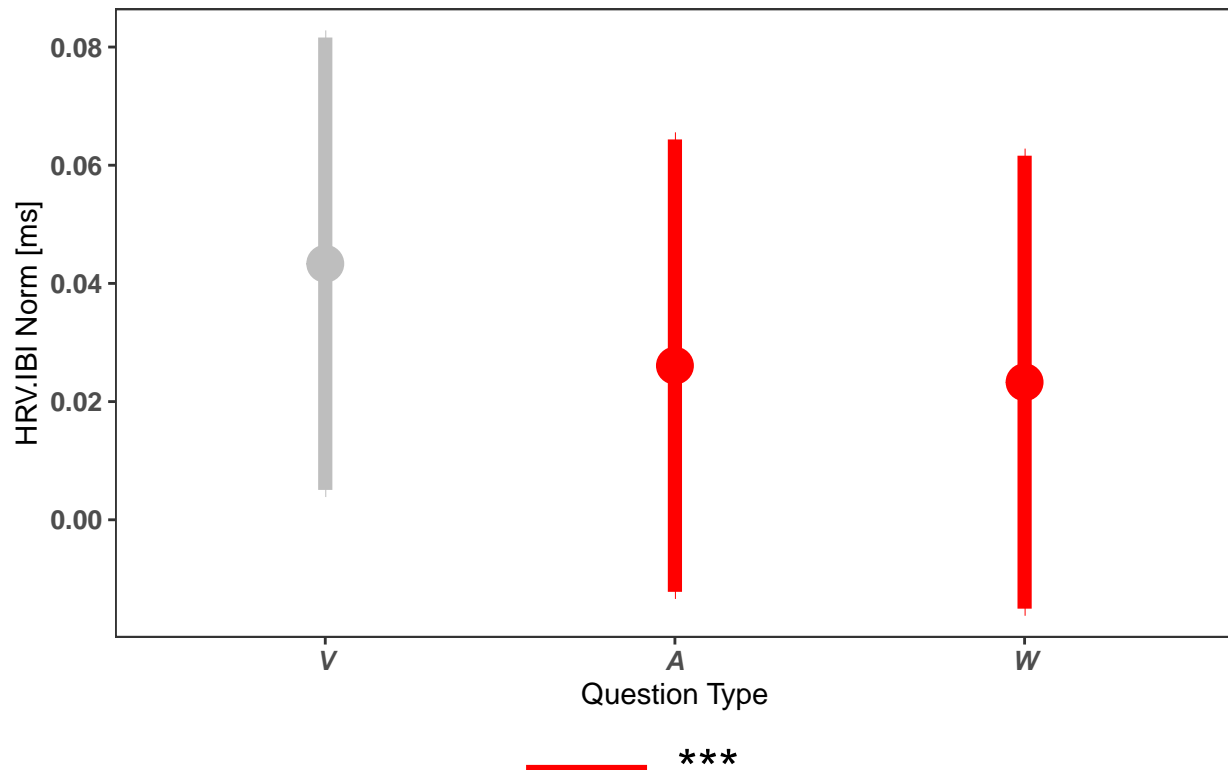
Random Effects Plot



We separately plotted the random effects from this plot we can observe that there is a lot of variability in coefficients ranging from -0.26 to 0.34 hence indicating the significance or importance considering interindividual variability in our model.

- Some participants like S016 and S013 show notably high positive deviations (0.34 and 0.15 respectively), indicating a higher than average response, possibly suggesting lower stress levels or better physiological resilience to the exam conditions.
- Participants such as S005 and S009 demonstrate substantial negative deviations (-0.26 and -0.16 respectively), which might indicate increased stress or a more significant physiological response to the exam settings compared to their peers.
- Several participants, such as S014 and S011, show deviations very close to zero, suggesting their responses align closely with the average participant's response in the study.

Predictors plot

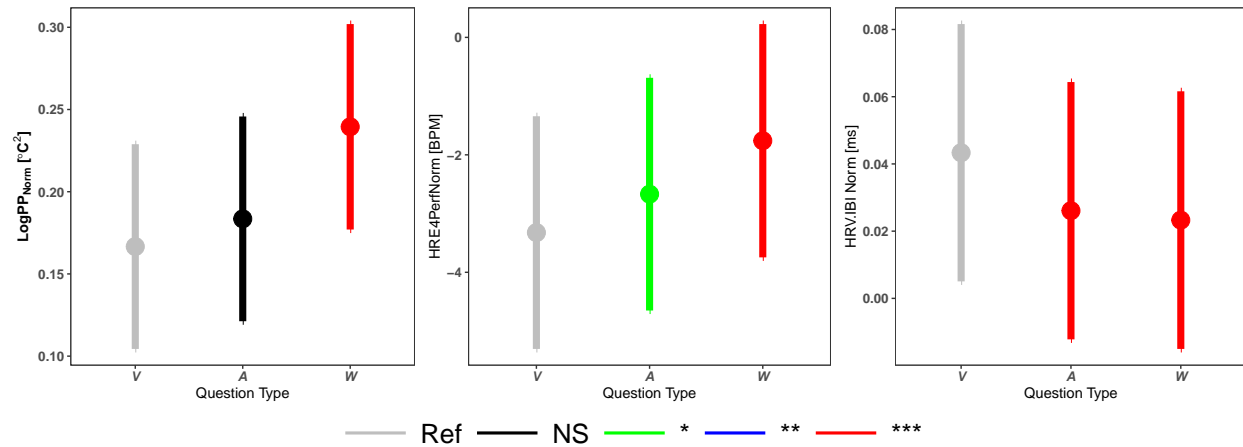


The above plot shows the normalized heart rate variability (HRV) responses to different question types in math exam:

- Abstract Questions (A) and Word Questions (W) both show significant decreases in HRV, marked by red and indicated with three asterisks (***), denoting a high level of statistical significance. This suggests that these question types are associated with increased cognitive or emotional stress, leading to reduced HRV, a marker generally associated with decreased parasympathetic nervous system activity and increased stress.

Considering these effects, we see from the predictors plot that moving from the reference video type to Abstract and Word type questions, HRV values decrease, indicating that students experience less stress when attempting humorous math video questions compared to the other types.

Comparison Plot for Three Stress Variables



From the analysis of the three model plots of stress indicators—perspiration (PP), heart rate (HR), and heart rate variability (HRV)—a consistent pattern emerges: participants show lower heart rate and perspiration levels when engaging with video type questions as compared to abstract and word type questions. Additionally, HRV values are notably higher for video type questions, suggesting that participants experience significantly lower stress levels with this format.

This contrasts sharply with the word type questions, which appear to be the most stressful, followed by abstract questions. Video questions, on the other hand, induce the least stress. This clear differentiation in physiological stress responses to different types of questions highlights the potential benefits of incorporating humorous video content in educational settings to alleviate stress and enhance learning experiences.

Step 4

Conduct logistic regression modeling using Accuracy.Score as the response variable. Justify your methodology and express your insights.

Full Model

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Accuracy.Score ~ QType + sQSolTime + sPPlogNorm + sHRE4PerfNorm +
## sHRVIBINorm + sSAI + (1 | QName)
##
##      AIC      BIC    logLik deviance df.resid
##    951.6    995.1   -466.8    933.6      918
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.0689 -0.5681  0.3070  0.5314  5.1327
##
## Random effects:
##  Groups Name            Variance Std.Dev.
##  QName (Intercept) 2.517      1.587
## Number of obs: 927, groups:  QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.01754    0.49341   0.036 0.971644
## QTypeA        1.46829    0.69745   2.105 0.035272 *
## QTypeW        1.28484    0.69741   1.842 0.065431 .
## sQSolTime     0.71501    0.11605   6.161 7.22e-10 ***
## sPPlogNorm    0.02765    0.08799   0.314 0.753348
## sHRE4PerfNorm 0.09387    0.08882   1.057 0.290575
## sHRVIBINorm   0.17287    0.09324   1.854 0.063748 .
## sSAI          -0.31681    0.08736  -3.626 0.000288 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) QTypeA QTypeW sQSlTm sPPlgN sHRE4P sHRVIB
## QTypeA        -0.718
## QTypeW        -0.716  0.520
## sQSolTime     -0.142  0.181  0.157
## sPPlogNorm    0.001  0.019 -0.018  0.129
## sHRE4PrfNrm   0.002  0.015 -0.011  0.067  0.002
## sHRVIBINorm  -0.010  0.026  0.022  0.074 -0.002  0.274
## sSAI          0.032 -0.049 -0.048 -0.220  0.051 -0.038 -0.228
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00267159 (tol = 0.002, component 1)
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
```

```

## Link function: logit
##
## MODEL FIT:
## AIC = 951.57, BIC = 995.06
## Pseudo-R2 (fixed effects) = 0.09
## Pseudo-R2 (total) = 0.48
##
## FIXED EFFECTS:
## -----
##           Est.    S.E.    z val.    p
## -----
## (Intercept)      0.02   0.49     0.04   0.97
## QTypeA           1.47   0.70     2.11   0.04
## QTypeW           1.28   0.70     1.84   0.07
## sQSolTime         0.72   0.12     6.16   0.00
## sPPlogNorm        0.03   0.09     0.31   0.75
## sHRE4PerfNorm     0.09   0.09     1.06   0.29
## sHRVIBINorm       0.17   0.09     1.85   0.06
## sSAI             -0.32   0.09    -3.63   0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group   Parameter   Std. Dev.
## -----
##   QName   (Intercept)   1.59
## -----
##
## Grouping variables:
## -----
##   Group   # groups   ICC
## -----
##   QName      36      0.43
## -----

```

The logistic regression model aimed to determine factors influencing exam Accuracy Score, employing variables like question type, question solving time, and various physiological indicators. Notably, the model indicated significant effects from the solving time and State Anxiety Index (SAI), with solving time positively impacting accuracy scores ($p < 0.001$), suggesting that longer solving times may correlate with higher accuracy. Conversely, higher SAI scores negatively affected accuracy ($p < 0.001$), implying that increased anxiety could decrease performance.

The fixed effects of question types A and W are also statistically significant, suggesting that these question types notably affect accuracy. The random effects, centered around different question names (QName), highlighted the variability in accuracy scores across different exam questions, underscoring the individual impact of specific questions on performance.

Overall, the model explained a moderate proportion of variance in accuracy scores (pseudo R^2 total = 0.48), indicating a fair fit. The convergence issues noted suggest potential overfitting or the need for model simplification or adjustment of parameters.

Manual Optimization:

Removing sPPlogNorm

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula: Accuracy.Score ~ QType + sQSolTime + sHRE4PerfNorm + sHRVIBINorm +
##         sSAI + (1 | QName)
##
##           AIC          BIC    logLik deviance df.resid
##        949.7        988.3   -466.8    933.7        919
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.0525 -0.5664  0.3059  0.5293  5.1185
##
## Random effects:
##   Groups Name          Variance Std.Dev.
##   QName (Intercept) 2.512      1.585
## Number of obs: 927, groups:  QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.01630    0.49298   0.033 0.973620
## QTypeA         1.46524    0.69674   2.103 0.035465 *
## QTypeW         1.28978    0.69670   1.851 0.064131 .
## sQSolTime      0.71054    0.11507   6.175 6.62e-10 ***
## sHRE4PerfNorm  0.09366    0.08880   1.055 0.291532
## sHRVIBINorm    0.17303    0.09315   1.857 0.063243 .
## sSAI          -0.31827    0.08725  -3.648 0.000265 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) QTypeA QTypeW sQSlTm sHRE4P sHRVIB
## QTypeA       -0.718
## QTypeW       -0.716  0.521
## sQSolTime    -0.144  0.180  0.161
## sHRE4PrfNrm  0.002  0.015 -0.011  0.066
## sHRVIBINorm -0.010  0.026  0.022  0.076  0.274
## sSAI         0.033 -0.050 -0.048 -0.229 -0.039 -0.228
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
## Link function: logit
##
## MODEL FIT:
## AIC = 949.67, BIC = 988.33
## Pseudo-R2 (fixed effects) = 0.09
## Pseudo-R2 (total) = 0.48
##
```

```
## FIXED EFFECTS:
## -----
##           Est.   S.E.   z val.   p
## -----
## (Intercept)      0.02   0.49    0.03   0.97
## QTypeA           1.47   0.70    2.10   0.04
## QTypeW           1.29   0.70    1.85   0.06
## sQSolTime        0.71   0.12    6.18   0.00
## sHRE4PerfNorm    0.09   0.09    1.05   0.29
## sHRVIBINorm      0.17   0.09    1.86   0.06
## sSAI             -0.32   0.09   -3.65   0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group   Parameter   Std. Dev.
## -----
##   QName   (Intercept)   1.58
## -----
##
## Grouping variables:
## -----
##   Group   # groups   ICC
## -----
##   QName      36      0.43
## -----
```

When we remove PPlogNorm, we observe a significant decrease in AIC and BIC values, while the model's explainability (pseudo R-squared) remains unaffected. From the resulting model summary, we can observe that the least significant predictor is sHRE4PerfNorm, so we consider removing it in the next step.

Removing sHRE4PerfNorm

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula: Accuracy.Score ~ QType + sQSolTime + sHRVIBINorm + sSAI + (1 |
##     QName)
##
##      AIC      BIC   logLik deviance df.resid
##   948.8    982.6  -467.4   934.8     920
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.0527 -0.5523  0.3057  0.5267  4.7745
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   QName   (Intercept) 2.492    1.578
## Number of obs: 927, groups:  QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.01532    0.49101   0.031  0.97511
## QTypeA       1.45622    0.69390   2.099  0.03585 *
```

```

## QTypeW      1.29951    0.69398    1.873    0.06113 .
## sQSolTime   0.70322    0.11451    6.141 8.21e-10 ***
## sHRVIBINorm 0.14546    0.09037    1.610 0.10749
## sSAI        -0.31500    0.08692   -3.624 0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) QTypeA QTypeW sQSlTm sHRVIB
## QTypeA      -0.718
## QTypeW      -0.716  0.521
## sQSolTime   -0.144  0.179  0.162
## sHRVIBINorm -0.012  0.023  0.027  0.064
## sSAI         0.033 -0.049 -0.048 -0.228 -0.230
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
## Link function: logit
##
## MODEL FIT:
## AIC = 948.76, BIC = 982.58
## Pseudo-R2 (fixed effects) = 0.09
## Pseudo-R2 (total) = 0.48
##
## FIXED EFFECTS:
## -----
##              Est.    S.E.    z val.    p
## -----
## (Intercept)    0.02    0.49     0.03    0.98
## QTypeA          1.46    0.69     2.10    0.04
## QTypeW          1.30    0.69     1.87    0.06
## sQSolTime       0.70    0.11     6.14    0.00
## sHRVIBINorm     0.15    0.09     1.61    0.11
## sSAI           -0.32    0.09    -3.62    0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group   Parameter   Std. Dev.
## -----
##   QName   (Intercept)   1.58
## -----
##
## Grouping variables:
## -----
##   Group   # groups   ICC
## -----
##   QName      36      0.43
## -----

```

After removing sHRE4PerfNorm, we can observe from the model summary that the AIC value decreased slightly compared to the previous model, whereas the BIC value decreased more significantly. The R-

squared value remains unaffected, indicating good model explainability even after removing sHRE4PerfNorm. Additionally, we notice that sHRVIBINorm is the least significant predictor, with a p-value close to 0.1. Therefore, we consider removing it in the next iteration.

Removing sHRVIBINorm

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Accuracy.Score ~ QType + sQSolTime + sSAI + (1 | QName)
##
##          AIC          BIC    logLik deviance df.resid
##       949.3       978.3   -468.7    937.3      921
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.8665 -0.5516  0.3110  0.5305  5.2235
##
## Random effects:
##  Groups Name            Variance Std.Dev.
##  QName (Intercept) 2.467      1.571
## Number of obs: 927, groups:  QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.02542    0.48866   0.052 0.958509
## QTypeA       1.43421    0.69038   2.077 0.037763 *
## QTypeW       1.27305    0.69042   1.844 0.065200 .
## sQSolTime     0.69345    0.11362   6.103 1.04e-09 ***
## sSAI          -0.28457    0.08448  -3.368 0.000756 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) QTypeA QTypeW sQSlTm
## QTypeA      -0.718
## QTypeW      -0.716  0.521
## sQSolTime   -0.144  0.178  0.161
## sSAI         0.032 -0.045 -0.044 -0.218
##
## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
## Link function: logit
##
## MODEL FIT:
## AIC = 949.33, BIC = 978.32
## Pseudo-R2 (fixed effects) = 0.08
## Pseudo-R2 (total) = 0.48
##
## FIXED EFFECTS:
## -----
##              Est.    S.E.    z val.    p
```



```
## -----
## (Intercept)      0.03  0.49    0.05  0.96
## QTypeA           1.43  0.69    2.08  0.04
## QTypeW           1.27  0.69    1.84  0.07
## sQSolTime        0.69  0.11    6.10  0.00
## sSAI             -0.28  0.08   -3.37  0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group      Parameter      Std. Dev.
## -----
##   QName      (Intercept)      1.57
## -----
##
## Grouping variables:
## -----
##   Group  # groups  ICC
## -----
##   QName      36      0.43
## -----
```

In this iteration, we eliminated the nonsignificant predictor sHRVIBINorm from the model. After reviewing the updated model summary and comparing it with the prior version, we noted a minor increase in the AIC by one point. However, there was a decrease of about four units in the BIC, and the model's explanatory power (pseudo R-squared) remained consistent. Although there was a small rise in AIC, the reduction in BIC and the stable pseudo R-squared support our decision to exclude sHRVIBINorm in the next iteration of the model. Additionally, while the question type remains a significant predictor, it is less impactful compared to the solving time for questions and the state anxiety scores. To further explore the model's explanatory potential, we removed the question type in the subsequent iteration.

Removing QType

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula: Accuracy.Score ~ sQSolTime + sSAI + (1 | QName)
##
##      AIC      BIC  logLik deviance df.resid
##   950.1    969.4  -471.1   942.1     923
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.7273 -0.5481  0.3123  0.5329  4.9311
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   QName (Intercept) 2.892    1.701
## Number of obs: 927, groups:  QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.94284    0.29898   3.154 0.001613 **
## sQSolTime    0.66199    0.11216   5.902 3.58e-09 ***
## sSAI         -0.27976    0.08438  -3.315 0.000915 ***
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) sQSlTm
## sQSolTime  0.029
## sSAI      -0.017 -0.217

## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
## Link function: logit
##
## MODEL FIT:
## AIC = 950.11, BIC = 969.43
## Pseudo-R2 (fixed effects) = 0.07
## Pseudo-R2 (total) = 0.50
##
## FIXED EFFECTS:
## -----
##              Est.   S.E.   z val.   p
## -----
## (Intercept)    0.94   0.30    3.15   0.00
## sQSolTime      0.66   0.11    5.90   0.00
## sSAI          -0.28   0.08   -3.32   0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group   Parameter   Std. Dev.
## -----
##   QName   (Intercept)   1.70
## -----
##
## Grouping variables:
## -----
##   Group   # groups   ICC
## -----
##   QName      36      0.47
## -----

```

After removing the question type, a clear improvement is evident when comparing the model summary to the previous iteration. Although the AIC increased very slightly (by one unit), the BIC reduced significantly (by approximately 10 units). Another interesting observation is that the model's explainability increased from a pseudo R-squared of 48% to 50%. Despite the slight increase in AIC, the decrease in BIC and the increase in model explainability outweigh it. Therefore, we plan to remove the question type from predictors. As a result, the final model obtained comprises only two highly significant predictors: question-solving time and State Anxiety Inventory scores.

Final Model

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Accuracy.Score ~ 1 + sQSolTime + sSAI + (1 | QName)
##
##      AIC      BIC    logLik deviance df.resid
##    950.1    969.4   -471.1    942.1      923
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.7273 -0.5481  0.3123  0.5329  4.9311
##
## Random effects:
##   Groups Name      Variance Std.Dev.
##   QName (Intercept) 2.892    1.701
## Number of obs: 927, groups: QName, 36
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.94284    0.29898   3.154 0.001613 **
## sQSolTime    0.66199    0.11216   5.902 3.58e-09 ***
## sSAI        -0.27976    0.08438  -3.315 0.000915 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) sQSlTm
## sQSolTime    0.029
## sSAI        -0.017 -0.217
##
## Analysis of Variance Table
##              npar Sum Sq Mean Sq F value
## sQSolTime     1 30.535  30.535  30.535
## sSAI          1 11.115  11.115  11.115
##
##
## Optimal Model of Manual Elimination AIC: 950.1065
## Full Model AIC: 951.5746
```

From the final model summary, it's evident that both question-solving time and State Anxiety Inventory questionnaire scores are highly significant predictors. Analyzing the coefficient estimates of these two predictors reveals the following:

Question-solving time: The coefficient associated with question-solving time has a positive effect on accuracy. This implies that accuracy tends to increase when participants take more time to solve the question. In other words, participants who spend more time on a question are likely to provide more accurate answers.

State Anxiety Inventory (SAI) scores: Conversely, the coefficient associated with SAI scores has a negative impact on accuracy. This suggests that students with higher SAI scores are more prone to providing less accurate answers compared to students with lower SAI scores. In essence, higher levels of state anxiety are associated with decreased accuracy in answering questions.

Comparison of Full and Optimal Models

Table 4: Compare the models for logistic regression modeling using Accuracy.Score as the response variable

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	Final Optimal Model HR	4	950.150	0.000	1.000	0.692	-471.053	0.692
1	Full Model for HR	9	951.771	1.621	0.445	0.308	-466.787	1.000

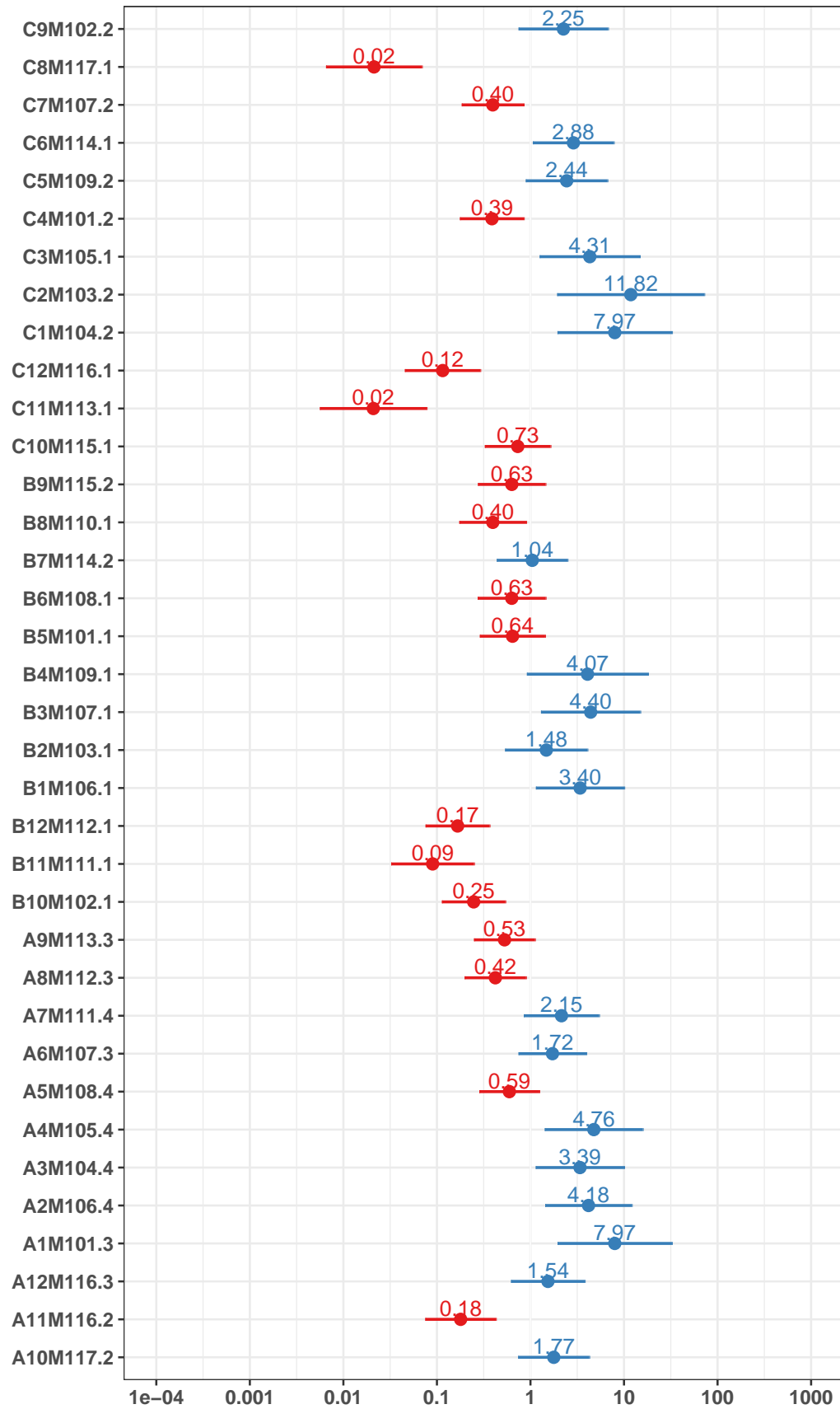
```
## MODEL INFO:
## Observations: 927
## Dependent Variable: Accuracy.Score
## Type: Mixed effects generalized linear regression
## Error Distribution: binomial
## Link function: logit
##
## MODEL FIT:
## AIC = 950.11, BIC = 969.43
## Pseudo-R2 (fixed effects) = 0.07
## Pseudo-R2 (total) = 0.50
##
## FIXED EFFECTS:
## -----
##               Est.   S.E.   z val.   p
## -----
## (Intercept)    0.94   0.30    3.15   0.00
## sQSolTime      0.66   0.11    5.90   0.00
## sSAI          -0.28   0.08   -3.32   0.00
## -----
##
## RANDOM EFFECTS:
## -----
##   Group   Parameter   Std. Dev.
## -----
##   QName   (Intercept)   1.70
## -----
##
## Grouping variables:
## -----
##   Group   # groups   ICC
## -----
##   QName      36      0.47
## -----
```

In comparing the full model and the final optimal model for logistic regression using Accuracy.Score as the response variable, the optimal model shows a more streamlined approach with fewer predictors (4 vs. 9 in the full model).

The AIC values indicate a minimal improvement (950.150 for the optimal model vs. 951.771 for the full model), suggesting a slightly better fit with fewer parameters. The AIC weight for the optimal model is higher (0.692 vs. 0.308), implying it is more likely to be the better model among the two compared.

This refinement effectively captures the significant predictors without overly complicating the model, which aligns with the principle of parsimony in statistical modeling achieving maximum explanatory power with the fewest possible predictors.

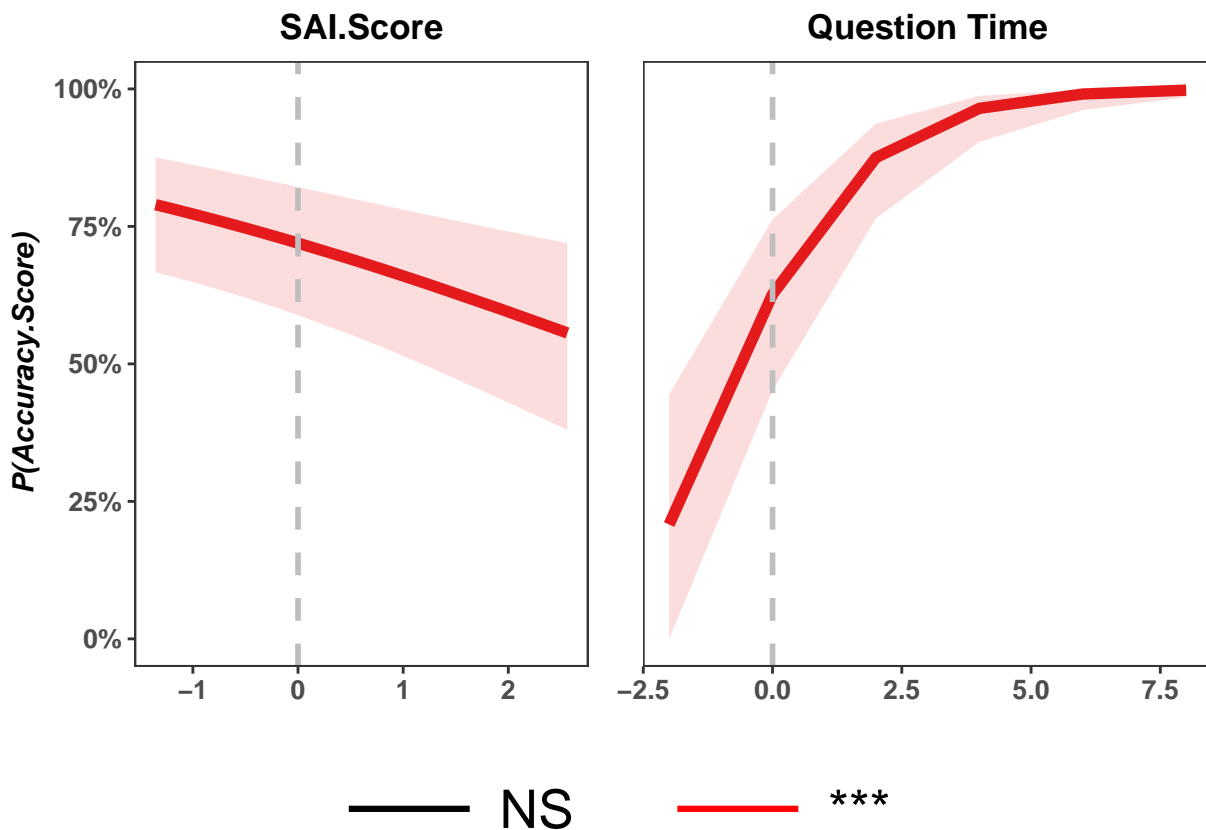
Random Effects Plot



The provided random effects plot showcases the variability of certain question types in terms of a specific outcome measure.

- The plot displays a range of effect sizes and confidence intervals for different questions, coded by type (V, A, W). Each horizontal line represents a question, with the dot indicating the effect estimate and the lines showing the 95% confidence interval.
- Question types 'V' (mostly on the left), 'A', and 'W' are scattered throughout the plot, showing different levels of influence. Notably, some questions like C3M105.1, A4M105.4, and C2M103.2 show particularly wide confidence intervals, suggesting less precision in the estimates for these questions.
- Several questions, particularly under type 'A' and 'W', exhibit more substantial effects (rightward on the x-axis) suggesting a stronger influence on the outcome measure. For instance, C2M103.2, A2M106.4, and B3M107.1 show significant positive effects.
- It's notable that while there's variability within each question type group, the 'A' and 'W' type questions tend to appear more frequently towards the right, suggesting these question types may generally elicit a stronger response or have a more substantial effect compared to 'V' type questions.

Predictors Plot



The analysis reveals the relationships between predictor variables, namely SAI.Score and Question Time, and the outcome variable, potentially representing the probability or log-odds of a correct response.

For SAI.Score, the left plot illustrates a nuanced pattern: initially, the predicted outcome rises slightly with increasing SAI.Score, yet beyond a certain threshold, it declines. This suggests a non-linear association or an optimal range for SAI.Score, where higher values may not necessarily lead to better outcomes.

Conversely, the right plot depicts Question Time's impact, showing a positive, sigmoidal relationship with the predicted outcome. This suggests that spending more time on a question tends to correlate with a higher

probability or log-odds of a correct response, albeit with diminishing returns at longer durations.

These visualizations provide insights into how SAI.Score and Question Time influence the outcome variable according to the final model's estimates.

Final Conclusions:

The findings indicate that:

- 1) W questions lead to higher arousal compared to V questions.
- 2) Pre-exam anxiety and time spent solving a question are crucial factors for exam performance, where lower anxiety and longer solving times are linked to better chances of giving correct answers.

Thus, presenting math exams in humorous video format (i.e., using V instead of W questions) would reduce stress without compromising performance