Evaluating the effect of emotions on academic achievement in mathematics

A quantile regression and Bayesian quantile regression approach

B.D.W. Janssen

Delft University of Technology

Evaluating the effect of emotions on academic achievement

A quantile regression and Bayesian quantile regression approach

by

B.D.W. Janssen

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Wednesday January 18, 2023 at 12:00 AM.

Student number:	5191343	
Project duration:	1 September, 2023 – January 18, 2023	
Thesis committee:	dr. A.J. Cabo,	TU Delft, supervisor
	dr. T. Janssen	TU Delft

TU Delft

Abstract

Quantile regression is a useful method to analyse data such that the estimates are more robust to outliers and the conditional distributions are more reliable for asymmetric distributions with respect to the commonly used ordinary least squares regression. Besides this, the quantile regression analysis might also include extra information on the conditional relations between the response variable and the explanatory variables. Therefore, it is previously used in educational sciences, among many other research areas. Bayesian statistics is an upcoming approach for computing estimates, as it allows prior knowledge modelling. The Bayesian quantile regression approach produces accurate parameter estimates by specifying prior distributions, likelihood estimators and MCMC methods to model an informative posterior distribution. In this research, the theory behind the quantile regression and the Bayesian quantile regression approach are considered. Especially Bayesian quantile regression for ordinal longitudinal data. This theory is then used on data of academic emotions to analyse its effect on attained grades of engineering students. Multiple aspects of quantile regression are included to analyse this effect, regarding gender, time and correlations between academic emotions. It was found that the quantile regression produced insights that were ignored by ordinary least squares regression, as the effects of anxiety altered over different quantiles. Especially when seperating genders, the effect of anxiety seemed to differ a lot between genders and different fractions of the response variable. Furthermore, an assumption for Bayesian quantile regression is made by specifying an exponentially distributed prior and by seperating the gender distributions, as was found by the estimates of the quantile regression approach.

B.D.W. Janssen
Delft, January 2023

Preface

This thesis is written as final requirement for the degree of Bachelor of Science in Applied Mathematics at Delft University of Technology, collaborating with PRogramme of Innovation in Mathematics Education (PRIME), which procured useful educational data for interesting research and goals.

I would like to thank Annoesjka Cabo for her involved guidance and advice during my bachelor project. The meetings were always very useful and I enjoyed sharing enthusiasm about the never ending applications and models.

B.D.W. Janssen
Delft, January 2023

Contents

1
Introduction

Do academic emotions have effect on the educational achievements of students? A rather difficult question. In this thesis the effects are analysed by different approaches. These effects are analysed on a deeper level by the introduction of quantile regression. Quantile regression allows us to see different effects for different student groups. Students with lower grades might experience a very different effect of emotions than students with higher grades. Therefore quantile regression has been used for educational sciences before.
Bayesian statistics has gained a lot of popularity in analysis as it allows the investigator to use previous knowledge and research to lead to a posterior distribution. Therefore also the Bayesian quantile regression approach is introduced.

1.1. Thesis outline

This thesis starts with an introduction to quantile regression to get an intuition of its use and importance. In Chapter 3 introduces the Bayesian quantile regression. First an introduction to Bayesian statistics and its significance in modern statistics is given, of which then the Bayesian method is applied to quantile regression. This includes information on calculating likelihood functions as well as choosing priors and sampling methods for Bayesian quantile regression. This method leads to an informative and reliable posterior distribution, which can be used in further research. Especially the Bayesian quantile regression for ordinal longitudinal data is elaborated in detail, as quantile regression had not been used for such data before the research of Alhamzawi and Ali [27] and as the research data is ordinal longitudinal. After the elaboration of the quantile regression approaches, a way of modelling quantile regression is introduced in chapter 4. Which includes analysis on how emotions influence the grades of students by using quantile regression on the researched data. Besides this research, the difference in gender and time are also analysed with the available data. Followed by a conclusion and discussion on how this research may be further conducted by modelling Bayesian quantile regression and using prior knowledge.

2

Quantile regression

Regression analysis generally investigates the relationship between variables. The variables that are considered are one quantitative response variable, say Y, which is dependent on one or more explanatory variables, X. The most common form of regression analysis is linear regression, where the response variable is a linear combination of the parameters. Therefore, say n is the number of observations and p is the amount of independent explanatory variables, a basic model for a multiple linear regression is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \epsilon_i \tag{2.1}$$

where $i = 1, \ldots, n$ is the i-th observation, β the vector of parameters that should be estimated and ϵ_i the error term, the so called residual.
So linear regression analysis examines the conditional distribution of the response variable Y as a function of several X's. Therefore regression analysis can be used to quantify the strength of the correlation of variables. However, the correlation should rather be interpreted as an estimate than a perfect prediction. The certainty of the estimate can be analysed by the residual. The bigger the residual, the less certain the information is about the correlation. However, when the estimates are considered to be reliable, this information can be used to predict or forecast. When the correlation between the Y and X variables are determined, the future Y data can be predicted by forming the same regression model on known independent explanatory variables. In this way regression analysis is used in research in biological, economic, social science, financial and behavioral fields, besides others.
A commonly used method to estimate the regression parameters is the least squares approach, which minimizes the squared distance between real data points and the fitted regression line. The quadratic loss function is defined as:

$$\sum_{i=1}^{n} (\beta^T \hat{x}_i - y_i)^2 \tag{2.2}$$

y_i being the actual response data points of each observation i, $\hat{x}_i$ is the vector of explanatory variables for each observation and $\hat{\beta}$ is the vector of unknown parameters. The vector parameter β for which this function is minimized is calculated by the following ordinary least squared formula:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} (\beta^T \hat{x}_i - y_i)^2 \tag{2.3}$$

By this minimization problem, the optimal β values can be estimated for k = 1, ..., p. The estimated parameters can be used in the multiple linear regression model introduced in equation 2.1 to plot a line such that the relationship between the response variable and the explanatory variables is shown. Most commonly the mean of the response values is used as affine function. As an example the data points in Figure 2.1 are examined. This is a univariate model with the Grade Point Avarage score as explanatory variable of student i ranging from 1.0 to 4.0. The response variable is the American College Testing (a standardized test before college) score on the y-axis corresponding to the same student, ranging from 1 to 36. The parameters are estimated by the ordinary least squared method to find the red line estimating the relationship between the variables. The blue lines represent the distance between the fitted line and the actual data points. From the data points it is suspected to have a positive relationship between the response variable and the explanatory variable, the OLS regression makes it possible to plot a line to show the relationship between the variables, which shows an obviously positive relationship.

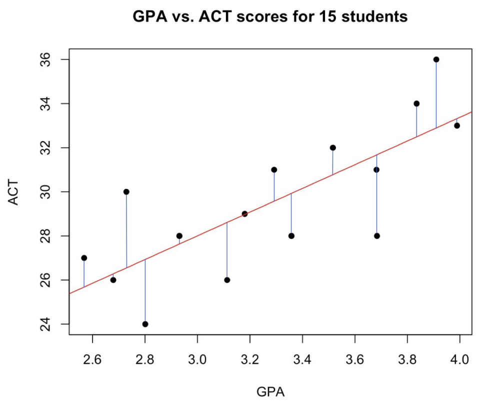

Figure 2.1: Least squares fit on 15 datapoints of students

As mentioned, such linear regression models can be useful to analyse the relationship between variables in data. However linear regression can have some shortcomings, such as its sensitivity to outliers, the assumption of independent covariates and its narrow information causing sensitivity to underfit. Therefore this chapter investigates the use of quantile regression. This is a form of regression analysis, which is more robust against outliers and skewed distributions than the common linear regression model. Besides that, quantile regression may discover predictive relationships between variables when there is a minimal relationship between the means of such variables. Therefore quantile regression can be used to analyse abnormal growth or connectivity. Since the research of Koenker and Bassett (1978)[17] on this model, quantile regression has become increasingly popular in several areas such as ecology, finance, social sciences and healthcare because of the stated aspects.

2.1. Quantile regression by minimization loss function

Quantile regression observes a predicted response at each quantile. Quantiles are points taken at regular intervals from the distribution function F. Therefore observing responses at different quantiles is as if the relationship between the variables is estimated for different interval groups within a distribution function F.

Definition 1 (Quantile)
F being the distribution function of a random variable Y and $\tau \in (0,1)$, *then y is the* τ*-th quantile if it satisfies:*

$$F(y) = \tau$$

For the τ-th quantile, 100*τ% of the response values is on the left side of the quantile and 100(1-τ)% on the right side of this quantile. For a probability distribution, the τ-th quantile specifies a value $Q(\tau)$ such that the probability that a random variable is on the left side of this value equals the given probability quantile τ. So in mathematical notation:

$$F_Y(Q_Y(\tau)) = P(Y \le Q_Y(\tau)) = \tau \tag{2.4}$$

Therefore the quantile function can be defined by taking the inverse of this distribution, leading to the following definition of the quantile function:

Definition 2 (τ-Quantile) *[9]*

$$Q_Y(\tau) = F_Y^{-1}(\tau) = inf\{y : F(y) \ge \tau\}, \qquad \tau \in (0,1). \tag{2.5}$$

An illustration of such quantile function can be seen in Fig 2.2. The probabilities τ can be found on the y-axis, the corresponding τ-th quantile $Q(\tau)$ can be found on the x-axis.

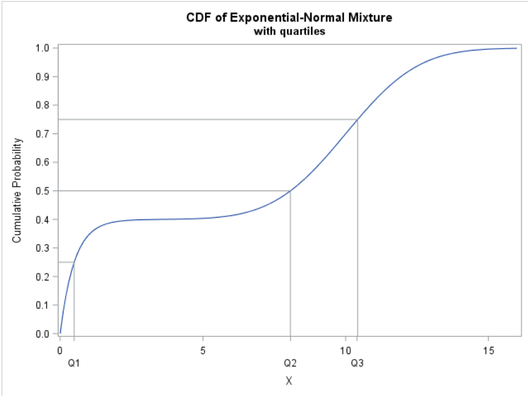

Figure 2.2: CDF of exponential-normal mixture wit Quantiles

A median is a quantile of the fraction $\tau = 0.5$. For the median the minimization of the sum of absolute residuals must equate the number of positive and negative residuals, such that there are the same number of observations on the left side of the median as on the right of the median. This is an example of the *Least Absolute Value* (LAV) regression, as the LAV minimizes the sum of absolute deviations. The LAV is unweighted in this example as the symmetry of the median is used. However for other quantiles the absolute residuals must be asymmetrically weighted. The weight is used to adjust the loss function so the relative importance of each quantile is taken into account. This can be seen as a natural generalization of L1-norm regression , to measure the regression relationship at several points of the distribution of y [17]. Therefore the weight is a function of the quantile level:

Definition 3 (Loss function)

$$\rho_\tau(u) = u(\tau - I_{(u<0)}), \tag{2.6}$$

$I_{(u<0)}$ being the indicator and u the difference between the real data points and the quantile regression line. The weight is τ to positive residuals and $(1-\tau)$ to negative residuals.

The loss function in this case, contrary to the quadratic loss function of mean regression in equation 2.2, minimizes the distance between the estimated quantile and real data quantile for the parameter estimation of the quantile regression. To find the τ-sample quantile regression, the expected loss function must be minimized. For random sample $\{y_1, y_2, ..., y_n\}$, the estimated quantile is defined by:

Definition 4 (Quantile estimator)

$$\hat{Q}(\tau) = \underset{Q\in\mathbb{R}}{arg\,min} \sum_{i=1}^{n} \rho_\tau(y_i - Q), \tag{2.7}$$

$$= \underset{Q\in\mathbb{R}}{arg\,min} \left[(\tau - 1) \sum_{y_i<Q} (y_i - Q) + \tau \sum_{y_i \geq Q} (y_i - Q) \right] \tag{2.8}$$

2.2. Conditional quantiles

For the conditional quantile regression, to estimate the quantile function, a parametric function of independent variables replaces the scalar Q in the quantile estimator in Definition 4. These variables are the explanatory variables of which the relationship with the response variable is analysed per quantile. Therefore the τ-th quantile is assumed to be a function of the independent variables. β_τ is defined as the vector of unknown quantile coefficients dependent on τ and X as the covariate vector that is evaluated at $\tau \in (0,1)$. giving the following conditional quantile random variable:

Definition 5 (τ-th conditional quantile of Y given X)
If $F_{Y_i|X_i}(Q_{Y|X_i}(\tau)) = \tau$ as in equation 2.4, the conditional quantile is the inverse cumulative distribution function of Y given at X_i. The τ-th conditional quantile of Y given X is:

$$Q_{Y_i|X_i}(\tau) = X_i^T \beta_\tau$$

Then given the distribution of Y, the values for the τth regression quantile are estimated with parameter estimations $\hat{\beta_\tau}$, again by minimizing the loss function:

Definition 6 (Parameter estimator)

$$\hat{\beta_\tau} = \underset{\beta\in\mathbb{R}^k}{arg\,min} \sum_{i=1}^{n} (\rho_\tau(Y_i - X_i\beta)). \tag{2.9}$$

The estimated β parameters that are drawn from this minimization problem illustrate the estimated effects of the independent variables on the dependent variable. In the *Ordinary Least Squared* linear regression, there is only information on the estimated mean parameter, generalising the effect of the independent variable. From equation 2.9, it can be seen that the estimated β is different for every quantile. Therefore the effect of the independent variable is inspected for different sets of the data.

2.3. Application of quantile regression

As stated earlier, a reason to use quantile regression instead of mean regression is that it is robust to outliers. An outlier can influence the mean value remarkably as every point is equally weighted. Therefore an outlier can have a big change on the mean, while it is just a single point in the data set. Contrarily, the median does not take into account the value of the data points, but rather looks at the place of the points in the distribution such that the probability of a datapoint is equally predicted to be on the left or right side of this median point. Therefore the value of an outlier does not have a big effect on the regression line formed by the median or other quantiles.
In figures 2.3 and 2.4 this is illustrated. The mean regression line has a notable change, only with the change of two single data points in a data set. The median is more robust to such outliers and does not differ too much with the first quantile regression line in figure 2.3.

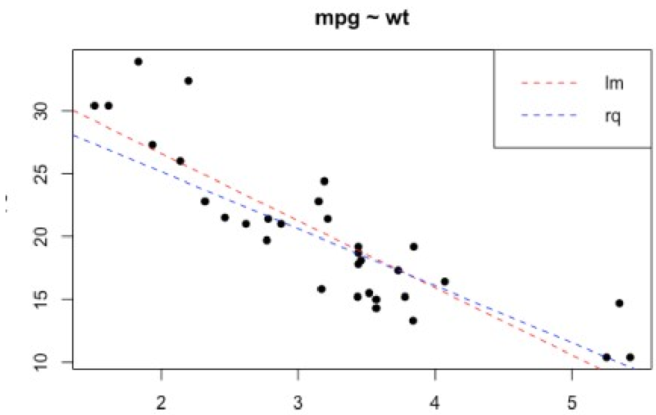

Figure 2.3: Linear Regression (lm) and Quantile Regression (rq) on dataset

Another example for which quantile regression gives more information than simple mean regression, is if change happens in the tails of the distribution instead of the distribution as a whole. For instance, a new route for a bus can decrease the mean walking distance for a customer to walk to the bus, as it shortens the distance a bit for 90% of the customers. However, the walking distance for 10% of the travelers can significantly decrease, especially if these are the customers that already had the longest walking distance to the bus station. A quantile regression line would show this significant unpleasant outcome, where it would not show in a mean regression line. Therefore, quantile regression is especially useful for asymmetric distributions of

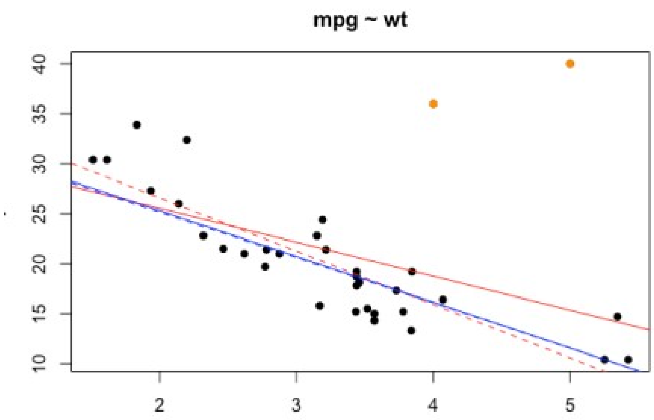

Figure 2.4: Linear Regression (lm) and Quantile Regression (rq) including Outliers

the response, as conditional effects are investigated in different parts of the distribution.

Quantile regression is also useful when heteroskedasticity occurs. Heteroskedasticity is the non-constant behaviour of standard deviations of a responsive variable. As quantile regression allows for a more detailed analysis of the relationship between the response variable and the predictors by focusing on specific quantiles of the distribution, it does not rely on the same distributional assumptions as ordinary least squares and thereby more robust to violations of the homoskedasticity assumptions. Furthermore, quantile regression can be used to identify the outliers and extreme data points, making it more effective in capturing non-linear relationships

2.4. Bootstrap method

The bootstrap method is often used for quantile regression, as it helps to estimate the uncertainty associated with the estimated quantiles. It is a statistical technique to estimate sampling distributions by repeatedly drawing samples from a dataset and creating a new data set with replacement. Therefore it is a useful tool for making inferences about a population from a smaller sample and thereby providing more accurate parameter estimates.
In Wild bootstrap (Wu and Liu, 1988)[6], the independent values stay on their initial value during the resampling. The resampling is based on the response variable and a modification of the residual values. Therefore each replicate y is computed by $y_i = \hat{y}_i + \hat{e}_i w_i$, where w_i is generated from a distribution as mentioned in The Wild Bootstrap procedure 2). To explain the wild bootstrap procedure, the steps of the algorithm are shown in The Wild bootstrap procedure algorithm below.

The Wild Bootstrap procedure

1) Fit a linear model to the data. $y_i = x_{1i}\beta_1 + \cdots + x_{ni}\beta_n + e_i$. Denote the estimate of the parameter vector by $\hat{\beta}$ and use $\hat{e}_i$ to represent the residuals.

2) Generate w_i from an appropriate distribution satisfying the condition $e_i^* = w_i \|\hat{e}_i\|$

3) Calculate the bootstrapped sample as $y_i^* = X_i^T \hat{\beta} + e_i^*$

4) Refit the linear model to the bootstrap sample and denote the bootstrap estimate by $\hat{\beta}^*$

5) Repeat steps 2-4 multiple times untill the requested amount of population observations are found. Say this is for repeating the procedure for B times, then estimate the variance of $\hat{\beta}$ by the sample variance of the B copies of $\hat{\beta}^*$ formed in step 4.

The confidence intervals are generated from the range of values obtained from the different samples generated by this Wild bootstrap procedure.

3

Bayesian Quantile Regression

As illustrated in the previous chapter, quantile regression is a very informative regression analysis to model conditional expectations. However, only the frequentist linear quantile regression method is considered. Introducing Bayesian statistics might provide more accurate estimates of the conditional quantiles by introducing prior knowledge of the distribution. Keming Yu and Rana A. Moyeed [14] proposed a Bayesian framework for quantile regression to estimate the conditional quantiles. This Bayesian Quantile Regression (BQR) is useful for nonlinear distributions and sparse data, as prior knowledge is included. Besides this, the BQR also provides an estimate of the uncertainty of the estimated quantiles, which is useful to indicate the reliability of the outcome.
Therefore in this chapter, an introduction to Bayesian statistics, the Bayesian quantile regression and especially Bayesian quantile regression for ordinal longitudinal data will be further elaborated.

3.1. Bayesian Statistics

In Bayesian statistics, prior information is combined with observed data to generate a posterior distribution of potential outcomes. In this way it applies probability to statistical problems and tries to preserve and refine uncertainty by modifying individual beliefs instead of eliminating uncertainty by estimates. The approach of determining the posterior distribution is based on Bayes' theorem on conditional probability, stating:

Theorem 1 (Bayes' Theorem) *[21]*
For any two events A and B with $0 < P(A) < 1$ *and* $P(B) > 0$,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

for determining the probability of an event A, given the occurance of another event B.

Hyperparameters are model parameters whose value is used to control the learning process, therefore they are chosen before the start of an algorithm and estimated without using the actual data. Using Bayes' Theorem and the hyperparameters θ of the prior distribution, the posterior distribution can be defined as follows:

Definition 7 (Posterior Distribution) *[21] [19]*
Let x denote the observed realisation of a (possibly multivariate) random variable X with density function $f(x|\theta)$. *Specifying a prior distribution with density function* $f(\theta)$, *allows us to compute the density function* $f(\theta|x)$ *of the* posterior distribution *using Bayes' theorem*

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)\, d\theta},$$

(For discrete parameters θ, *the integral in the denominator has to be replaced with a sum)*

This leads to a density function that is conditional on prior knowledge and includes the density function of the observed realisation. This idea of working with data and prior knowledge for predicting estimates is also used in quantile regression, as is described in the following section.

3.2. Bayesian quantile regression

To illustrate the use of Bayesian statistics to estimate useful and reliable parameters for the quantile regression, the structure of this approach is illustrated in Algorithm 1.

Algorithm 1 Bayesian Quantile Regression

1. Specify a prior for the regression coefficients and hyperparameters. Priors used for Bayesian Quantile Regression are elaborated in subsection 3.2.2.

2. Calculate the likelihood of the data given the prior. The likelihood function in quantile regression is based on the Asymmetric Laplace Distribution as described in subsection 3.2.1.

3. Use Markov chain Monte Carlo (MCMC) to generate samples from the posterior distribution of the regression coefficients and hyperparameters, as illustrated in subsection 3.2.3.

4. Calculate the posterior mean and variance of the regression coefficients and parameters and calculate the quantile regression predictions.

The steps of this algorithm are further elaborated in the following subsections.

3.2.1. Bayesian inference by forming a likelihood function based on asymmetric Laplace distribution

In the linear regression approach, the quantile distribution is determined by the minimization of the loss function. The minimization of the loss function is equivalent to the maximization of the likelihood formed by combining independently distributed Asymmetric Laplace densities[14] [29], the derivation of this is shown below. Therefore, Bayesian inference for quantile regression proceeds by forming the likelihood function based on the *Asymmetric Laplace Distribution* (ALD), regardless of the actual distribution of the data that is analysed[17][14].

The ALD is characterized by:

Definition 8
A random variable X has an Asymmetric Laplace Distribution with parameters m, λ, k, noted as ALD(m, λ, k), when its probability density function is:

$$f(x; m, \lambda, k) = \left(\frac{\lambda}{k + \frac{1}{k}}\right) e^{-(x-m)\lambda \; sgn(x-m) k^{sgn(x-m)}} \tag{3.1}$$

With parameters:

$$\begin{aligned} m &= \texttt{location parameter} \\ \lambda > 0 &= \texttt{scale parameter} \\ k > 0 &= \texttt{asymmetry parameter} \end{aligned}$$

To use this distribution in the quantile regression context, it is useful to use a parametrization of the mean and variance of the *Asymmetric Laplace Distribution* (ALD) with the percentile parameter τ involved. The parametrization parameters are therefore defined as:

$$\mu = \qquad m + \frac{1-2\tau}{\tau(1-\tau)\lambda}$$

$$\sigma^2 = \qquad \frac{1-2\tau+2\tau^2}{\tau(1-\tau)}\lambda^2$$

Concluding the new parameterized probability density function of the ALD:

$$f_\tau(u; \sigma, \mu) = \frac{\tau(1-\tau)}{\sigma} \begin{cases} exp(-((\tau-1)/\sigma)(u-\mu)) & \text{if } u \leq \mu \\ exp(-(\tau/\sigma)(u-\mu)) & \text{if } u > \mu \end{cases} \tag{3.2}$$

Now using the function introduced in equation 2.6, the following function can be derived:

$$\rho_\tau\Big(\frac{u-\mu}{\sigma}\Big) = \frac{u-\mu}{\sigma}\Big(\tau - I_{(u-\mu<0)}\Big) = \begin{cases} -((\tau-1)/\sigma)(u-\mu) & \text{if } u \le \mu \\ -(\tau/\sigma)(u-\mu) & \text{if } u > \mu \end{cases} \tag{3.3}$$

Therefore the density of the asymmetric Laplace function can be rewritten as:

$$f_\tau(u;\sigma,\mu) = \frac{\tau(1-\tau)}{\sigma} e^{-\rho_\tau(\frac{u-\mu}{\sigma})} \tag{3.4}$$

The likelihood function is used to estimate parameters by indicating which parameter values make the observed data more probable. Therefore the maximization of the likelihood function estimates under which statistical model the observed data is most probable.
Assuming that $y_i \sim ALD(\sigma,\mu,\tau)$ and location parameter $\mu_i = x_i^{'}\beta$ and σ a nuisance parameter, by the Neyman-Pearson Lemma, the likelihood for n independent observations can be formed.

Lemma 1 (Neyman-Pearson Lemma)
If $X_1,\ldots,X_n$ is a random sample of size n from a distribution with probability desity function $f(y;\theta)$, then the joint probability density function of $X_1,\ldots,X_n$ is denoted by the likelihood function:

$$L(\theta) = L(\theta;x_1,\ldots,x_n) = f(x_1;\theta)\times\cdots\times f(x_n;\theta)$$

Therefore using this lemma, the likelihood function of β can be generated such that:

$$L(\beta,\sigma;y,\tau) = \tau^n(1-\tau)^n exp\Big\{-\sum_{i=1}^{n}\rho_\tau(y_i - x_i^{'}\beta)\Big\} \tag{3.5}$$

The maximization in the likelihood of equation 3.5 with respect to parameter β is then equivalent to the minimization of the loss function described in Chapter 2.

3.2.2. Priors

In Bayesian statistics, a conjugate prior is one that is in the same probability distribution family as the posterior. The parameters of the distribution is then determined by any existing belief or information[4].
However, for the quantile regression formulation of Bayesian statistics, there is not a standard conjugate prior distribution[14]. Despite this, the MCMC methods described in subsection 3.2.3 allows the use of any prior distributions to form a posterior distibution. Therefore any available information could form a prior distribution for the bayesian quantile function. However, in practise the use of an improper uniform prior distribution is common. A prior $\pi(\beta)$ is said to be improper if $\int\pi(\beta) = \infty$. Therefore indeed the uniform prior $\pi(\beta) \propto 1$ is clearly improper and noninformative. Therefore this uniform prior distribution is often used, as the result of the joint posterior distribution is proportional to the likelihood. For this we can use the theorem of Yu and Moyeed (2001) stating:

Theorem 2 (Improper priors for parameters)
If the likelihood function is equal to $L(\beta,\lambda;y,\tau) = \tau^n(1-\tau)^n exp\Big\{-\sum_{i=1}^{n}\rho_\tau(y_i - x_i^{'}\beta)\Big\}$ and the prior, p(β) = 1, then the posterior distribution of β, $\pi(\beta)$, will have a proper distribution. Therefore,

$$0 < \int \pi(\beta|y)d\beta < \infty.$$

–
Given the observations$y = (y_1,\ldots,y_n)$, the posterior distribution of β is then given by: [14]

$$\pi(\beta|y) \propto L(y|\beta)p(\beta) \tag{3.6}$$

$\propto$ meaning proportional to.

Despite the improper priors can be used, the inference on the data could be more reliable if priors are based on historical data. Besides this, it may be more reasonable to have different priors for different quantiles instead of having the same parameter values for modelling the quantiles. Therefore, Keming Yu and Rahim Alhamzawi (2011)[26] defined Power Priors for Bayesian Quantile Regression when historical data is available. The basic formulation of the power prior by Ibrahim and Chen [13] is specified as follows:

Definition 9 (Power Prior)
Let current study response data be denoted by D, the corresponding likelihood function $L(\theta|D)$ with θ indicating the parameters. Then suppose historical data D_0 is available from a previous study, denote $L(\theta|D_0)$ as the likelihood function of the previous study. Then the Power prior is denoted as $\pi(\theta|D_0, a_0)$ and is proportional to:

$$\pi(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta|c_0),$$

$0 \leq a_0 \leq 1$ *being a scalar parameter and c_0 being a hyperparameter of the initial (improper) prior.*

In this formulation a_0 can be interpreted as a precision parameter of the historical data. Therefore when $a_0 = 0$ the power prior does not depend on the historical data at all. For $a_0 = 1$ the power prior corresponds to the posterior distribution of θ for historical data. The influence of the historical data on the final posterior can therefore be controlled by determining the value for a_0. However a proper prior distribution for a_0 can also be specified, such that:

$$\pi(\theta, a_0|D_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta|c_0)\pi(a_0|\delta_0).$$

Therefore the posterior distribution of the parameter θ using the power prior, as it is a function of the formulated power prior and the likelihood as calculated in previous section, is defined as:[13]

$$\pi(\theta|D, D_0, a_0) \propto L(\theta|D)L(\theta|D_0)^{a_0}\pi_0(\theta|c_0) \tag{3.7}$$

or using a proper prior for a_0 on hyperparameter δ_0:

$$\pi(\theta, a_0|D, D_0) \propto L(\theta|D)L(\theta|D_0)^{a_0}\pi_0(\theta|c_0)\pi(a_0|\delta_0) \tag{3.8}$$

The useful informative prior in Bayesian analysis by Ibrahim and Chen [13]is proven to be a proper joint prior for Bayesian quantile regression by Keming Yu and Rahim Alhamzawi [26]by the following statement:

Theorem 3 (Proper joint prior distribution in Bayesian quantile regression) *Suppose the initial prior distribution for θ is a uniform prior and a_0 has a beta prior. Then the joint prior distribution in quantile regression for (θ_τ, a_0), τ being the percentile, is proper:*

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^1 L(\theta_\tau|D_0)^{a_0} d a_0 d\theta_\tau < \infty$$

This statement is derived by the use of the theorem of the improper priors in Theorem 2 by Yu and Moyeed (2011) [14].

3.2.3. MCMC method obtaining a proper posterior distribution

The Markov chain Monte Carlo (MCMC) method can be implemented to extract posterior distributions for unknown parameters[8]. By the MCMC methods, the posterior characteristics will be estimated by obtained samples. These samples are extracted from a Markov chain which converges to the posterior distribution. A useful property of the Markov chain is that the probability of going from one sample to another only depends on the current state and not on the sequence of previous states. In mathematical notation:

Definition 10 (Markov Property) _
A Markov Chain is a sequence of random variables such that:

$$P(X_{t+1} = s|X_t = s_t, X_{t-1} = s_{t-1}, \dots, X_0 = s_0) = P(X_{t+1} = s|X_t = s_t).$$

So the idea is to develop a Markov Chain such that the stationary distribution converges to the distribution of the target posterior distribution. These chains move around randomly according to an algorithm looking for places with reasonably high contribution to the integral over the variable, assigning them higher probabilities. Then from the point that the Markov Chain is in the stationary distribution, drawing a sample from the chain is actually equivalent to sampling from the posterior distribution. The Monte-Carlo approach is then gaining information of the large sample instead of the distribution's equations[8].
Therefore Bayesian inference uses the information provided by the observed data in form of a likelihood function and a prior state about the parameters, to become a posterior state of beliefs. Then the MCMC method

draws a sequence of samples from the posterior. Then this becomes the next sample in the MCMC chain.

There are several sampling methods for this model. Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution. Therefore it is only useful in case of a multivariate distribution when sampling from the joint distribution is difficult, while the conditions are easy to sample. So it works by iteratively sampling from the full conditional distributions of each of the variables in the model, which is then used to update the values of the other variables[25][9]. Therefore the sequence of samples drawn from the jumping distribution is a random walk, which is considered to move around slowly. However, the Gibbs sampling is introduced as it is useful for the ordinal quantile regression model with longitudinal data as there are many variables that need to be updated. Therefore the constructing by Gibbs sampling of the conditional distributions is very useful to form an efficient posterior computation.[25] The sampler is known to converge to a unique invariant joint distribution, which is defined as the posterior distribution.
The Algorithm 2 shows the Gibbs Sampling.

Algorithm 2 Gibbs Sampling

1. Start by initializing values for all unknown parameters in the model: (x^0, y^0).

2. Change one parameter and keep the other parameter values fixed. Then sample from the conditional distribution of the changed variable: Sample $x^1 \sim p(x^1|y^0)$.

3. Now sample from the conditional distribution of the other variables, given the new fixed parameter value that is just sampled: $y^1 \sim p(y^1|x^1)$

4. Update the value for the parameters and repeat steps 2-3 untill convergence is reached. For the Gibbs sampler convergence is reached in the form of a unique invariant joint distribution. Then the output is the converged values as the approximate parameters of the target posterior distribution.

3.3. Bayesian Quantile Regression for Ordinal Longitudinal data

In the research of PRIME, ordinal longitudinal data is collected. Ordinal data meaning the data is classified in ranking, however the distance in value between these categories are unknown. The data is longitudinal as the same variables, students, are observed in different time periods. Longitudinal studies are often used for observations of thoughts, behaviors and emotions as the differences that can occur in the results are less likely to be the result of other external elements like cultural or age differences.
Quantile regression is not an obvious choice for ordinal response variables as it does not yield continuous quantiles that can be modeled via regression. However, Alhamzawi and Taha Mohammad Ali [25] adopted a Bayesian method for ordinal quantile regression using latent variable inferential framework of Albert and Chib [1] and the asymmetric Laplace distribution. Before this research in Bayesian statistics, ordinal longitudinal data had not been approached for estimating parameters for quantile regression.

3.3.1. Modeling ordinal longitudinal data for quantile regression approach

For modelling ordinal longitudinal data, the use of latent variables is very useful. Latent variables are unobserved response variables that are formed by information of other observable independent variables. This is advantageous for ordinal data expressing attitudinal statements with response alternatives like "strongly disagree" to "strongly agree" [28]. Therefore for the PRIME research, as mental states are difficult to observe, the latent variable is useful to translate the abstract concepts. The approach for analyzing the ordinal observed variables is by generating underlying normally distributed continuous variables [28]. This method is derived in this section.
The independent variables form the continuous latent random variable l_{ij} for student i at time j, are defined by:

$$l_{ij} = \alpha_i + x_{ij}^{'}\beta + \varepsilon_{ij},$$

α_i being the location-shift random effect. The location-shift random effect captures how different the effect is in time. This allows the effects to be different in each time period. The location-shift random effects are not estimated from the model but assumed to follow a multivariate normal distribution with mean zero and estimated covariance matrix, such that $\alpha_i \sim N(0,\phi)$.
As in the frequentist quantile regression model, x_{ij} represents the covariates with the corresponding β parameter. The ε_{ij} is a Skewed Laplace Distribution (SLD), as suggested by Geraci and Bottai (2007) [10], such that:

$$f(\varepsilon|\tau) = \tau(1-\tau)exp\Big\{-\rho_\tau(\varepsilon)\Big\}, \tag{3.9}$$

The unobserved latent response variable l_{ij} can then be used to assume the relationship with the actual observed ordinal data response defined by y_{ij}. The assumed relation between the latent response variable and the data response variable is as follows:

$$y_{ij} = \begin{cases} 1 & \text{if } \delta_0 < l_{ij} \leq \delta_1; \\ c & \text{if } \delta_{c-1} < l_{ij} \leq \delta_c; \\ C & \text{if } \delta_{C-1} < l_{ij} \leq \delta_C; \end{cases} \tag{3.10}$$

where $\delta_0, \ldots, \delta_C$ are points that define the bounds of intervals containing the observed outcome c such that the lower bound is δ_{c-1} and the upper bound is δ_c. Therefore the points satisfy $-\infty = \delta_0 < \delta_1 < \cdots < \delta_{C-1} < \delta_C = +\infty$ and $c = 2,\ldots,C-1$. This is useful as the observed data in y_{ij} is ordinal, therefore the calculated latent points are classified in groups corresponding to the ordinal points.

Then to find the CDF of y_{ij}, the information of the latent variable can be used. For that, it is useful to rewrite the SLD of ε such that:

$$\varepsilon_{ij} = (1-2\tau)v_{ij} + \sqrt{2v_{ij}}\epsilon_{ij}$$

as introduced by Kozumi and Kobayashi (2011) for Gibbs sampling methods[18]. The new latent variable that is introduced, v_{ij} now follows an exponential distribution with parameter $\tau(1-\tau)$. The ϵ_{ij} follows the standard normal distribution. Therefore the CDF of y_{ij} per ordinal response category can be rewritten as:

[25]

$$P(y_{ij} \le c|l_{ij},\delta_c) = P(l_{ij} \le \delta_c|\beta,\alpha_i,v_{ij}), \tag{3.11}$$
$$= P\Big(\alpha_i + x_{ij}^{'}\beta + (1-2\tau)v_{ij} + \sqrt{2v_{ij}}\epsilon_{ij} \le \delta_c\Big), \tag{3.12}$$
$$= P\Big(\epsilon \le \frac{\delta_c - \alpha_i - x_{ij}^{'}\beta - (1-2\tau)v_{ij}}{\sqrt{2v_{ij}}}\Big) \tag{3.13}$$

Therefore, as the distribution of ϵ is defined as the standard normal distribution, it can be concluded that:

$$P(y_{ij} \le c|l_{ij},\delta_c) = \phi\Big(\frac{\delta_c - \alpha_i - x_{ij}^{'}\beta - (1-2\tau)v_{ij}}{\sqrt{2v_{ij}}}\Big) \tag{3.14}$$
$$P(y_{ij} = c|l_{ij},\delta_{c-1},\delta_c) = P(\delta_{c-1} < l_{ij} \le \delta_c|\beta,\alpha_i,v_{ij}), \tag{3.15}$$
$$= \phi\Big(\frac{\delta_c - \alpha_i - x_{ij}^{'}\beta - (1-2\tau)v_{ij}}{\sqrt{2v_{ij}}}\Big) - \phi\Big(\frac{\delta_{c-1} - \alpha_i - x_{ij}^{'}\beta - (1-2\tau)v_{ij}}{\sqrt{2v_{ij}}}\Big) \tag{3.16}$$

with ϕ being the standard normal distribution.
Therefore, this can be seen as the response probability function, where every response category is effectively accounted for. Then for every data point, it can be elaborated how probable it is to happen.

3.3.2. Priors for ordinal longitudinal data

As mentioned in the subsection 3.2.2, any prior can be used to estimate Bayesian quantile regression parameters. Therefore, the zero-mean normal prior is usually used. However, when there are big differences in the size of fixed effects as in ordinal data, the zero-mean normal prior distribution performs poorly.[25]
Therefore, the Laplace prior distribution, which is a scale mixture of normals with an exponential mixing density, formed by Andrews and Mallows (1974) [2] is introduced:

$$P(\tau|\sigma) = \prod_{k=1}^{p} \frac{\sigma}{2} e^{-\sigma|\tau_k|} \tag{3.17}$$
$$= \prod_{k=1}^{p} \int_0^\infty N(\tau;0,s_k)Exp\Big(s_k;\frac{\tau^2}{2}\Big)ds_k \tag{3.18}$$

Therefore it can be concluded that for each β_k the zero-mean normal prior distribution with unknown variance is assigned. The exponential distribution has a parameter $\frac{\lambda^2}{2}$ for the variance, assuming it is independent.
For the cut-points δ introduced in the previous section, the order statistics are considered from the uniform $U(\delta_{min},\delta_{max})$, such that:

$$P(\delta) = (C-1)!\Big(\frac{1}{\delta_{max} - \delta_{min}}\Big)^{C-1} I(\delta \in T),$$

with $\delta = (\delta_0,\dots,\delta_C)$ and $T = \{(\delta_{min},...,\delta_{max})|\delta_{min} < \delta_1 < \cdots < \delta_{C-1} < \delta_{max}\}$.
In previous section, $P(\mathbf{y}|\mathbf{l},\delta)$ is defined. Furthermore $l_{ij}|v_{ij} \sim N(\alpha_i + x_{ij}^{'}\beta + (1-2\tau)v_{ij}, 2v_{ij})$. Now as the prior for the parameter β is defined in 3.17, the posterior distribution of all parameters and latent variables can be concluded to:

$$P(\beta,\alpha,\mathbf{l},\delta,\mathbf{v},\mathbf{s},\lambda^2,\phi|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{l},\delta)P(\mathbf{l}|\beta,\alpha,\mathbf{v})P(\delta)P(\mathbf{v}) \times P(\beta|\mathbf{s})P(s|\lambda^2)P(\alpha|\phi)P(\phi), \tag{3.19}$$

3.3.3. Gibbs Sampling

For an efficient posterior computation of the data, the Gibbs sampler is a useful method as it uses the conditional distributions. From the previous section, it can be concluded that the quantile regression posterior distribution includes a lot of conditional distributions. Therefore, this sampling algorithm is chosen above others. In Alhamzawi and Mohammad Ali (2018) [14], the algorithm for finding all parameter values is stated as follows:

Algorithm 3 Gibbs sampler for the Bayesian quantile regression posterior

1. Start with initial values for $[\beta^0, \alpha^0, l^0, \delta^0, v^0, s^0, \lambda^0, \phi^0]$

2. Update the initial values in order:
a. Generate v^t from $P(v|l^{t-1}, \beta^{t-1}, \alpha^{t-1})$
b. Generate β^t from $P(\beta|l^{t-1}, v^t, \alpha^{t-1}, s^{s-1})$
c. Generate s^t from $P(s|\lambda^{2(t-1)}, \beta^t)$
d. Generate $\lambda^{2(t)}$ from $P(\lambda^2|s^t)$
e. Generate α^t from $P(\alpha|l^{t-1}, \beta^t, v^t, \phi^{t-1})$
f. Generate ϕ^t from $P(\phi|\alpha^t)$
g. Generate l^t from $P(l|\beta^t, v^t, \alpha^t, \delta^{t-1})$
h. Generate δ^t from $P(\delta|y, l^t)$

3. Update the value for the parameters and repeat steps 2 untill convergence is reached. Then the output is the converged values as the approximate parameters of the target posterior distribution.

4

Application to PRIME research

In the article "Do Our Means of Inquiry Match our Intentions?" by Yaacov Petcher [24], quantile regression is applied to analyse the relationship between the well-being of children and their reading achievement in primary education. Quantile regression is a special case of conditional median modeling; the relationship between the grades and well-being of a child is researched more thoroughly than just by normal ordinary regression. Since the effects of the independent variables are explained on different quantiles of the dependent variable instead of using the mean, these give a more detailed explanation of this effect.
The following section includes research on education by a quantile regression approach comparrable to the research of Petscher [24]. However, this time the relationship between emotions of engineering students on a Mathematics course and the attained grades of these students is investigated. As emotions are difficult to test and challenging to rephrase into mathematical values, the *Achievement Emotions Questionnaire* (AEQ)[27] was used to extract useful data on emotions of the engineering students. This extraction of data was done by *The PRogramme of Innovation in Mathematics Education* (PRIME)[7]. Further introduction to this programme besides the introduction to the AEQ questionaire are established in this chapter followed by the quantile regression analysis. The quantile regression analysis gave a more detailed description of the data than normal mean regression, including results on gender differences, time differences and multi- versus univariate models of the emotions. The gender differences seemed to include new insights on the data regarding the effect of emotions on attained grades. Lastly, the chapter includes further results and conclusions of the quantile regression approach.

4.1. PRIME research

The *Programme of Innovation in Mathematics Education* (PRIME)[7] is a programme responsible for redesigning mathematics courses for TU Delft engineering students. It is active at TU Delft as part of the *Interfaculty Teaching of the department of Applied Mathematics* (DIAM). The intention of PRIME is to improve the mathematics courses in three different aspects; study results, the connection between mathematics and engineering and the active participation and motivation of students.
To improve mathematics courses, PRIME is researching the emotions, behaviour and motivations of students during their study of the mathematics course. In a recent study this data of engineering students was collected. This data is useful to inspect the relations between these emotions, behaviour and motivations and the obtained grades of the same students. Since for this research it is important to find out for whom the innovation in education is of positive effect, quantile regression can be used to analyse certain groups within the data group.

4.2. The Achievement Emotions Questionnaire

Emotions, behaviour and motivations are very hard to examine. However, PRIME asked engineering students from the TU Delft following the "Mathematics 1" course to answer a questionnaire to extract information of these students on these topics. The questionnaire investigated emotions, behaviour and motivations.The section on emotions was based on The Achievement Emotions Questionaire (AEQ)[27]. The AEQ is a validated multidimensional self-report instrument designed to assess college students' achievement emotions,

developed by two research grants from the German Research Foundation.
The AEQ assesses multiple emotions that are classified by the following four categories:

Positive Activating	Enjoyment, Hope, Pride
Positive Deactivating	Relief
Negative Activating	Anger, Anxiety, Shame
Negative Deactivating	Hopelessness, Boredom

The AEQ includes class-related, learning-related and test-related emotions, to thoroughly research the emotions of a student on a particular subject. These emotions of a student are scaled from 1 to 5, classified by the self-reported "strongly disagree"(1) to "strongly agree"(5) of the students to questions related to the multiple emotions. The questions measure the affective, cognitive, motivational and physiological components of each emotion. The AEQ is advised to be answered voluntarily for a more reliable outcome.
In the questionnaire of PRIME, the emotions of Relief and Hope were excluded from the questionnaire. Furthermore the questionnaire of PRIME included information of the students regarding gender, secondary school mathematics grade, studies and time spent on the mathematics course. The full questionnaire used by PRIME can be found in Appendix B .

4.3. Research Question

As the intention of PRIME is to improve the mathematical education for engineering students, it is useful to research whether the achieved scores of the students of the "Mathematics 1" course are related to the emotions of the students taking this course. Therefore the main research question is formulated as:

Do academic emotions on mathematics courses of engineers have an effect on their grade attained?

To answer this question, we want to analyse the following subquestions:

Is there a difference between certain groups (quantiles) of the data regarding the effect of academic emotions on the achieved grade?

Is there a difference between genders regarding the effect of academic emotions on grades?

Is there a difference in the long-run on the effect of the relationship between academic emotions and grades?

4.4. Hypothesis

To answer the first subquestion, *"Is there a difference between certain groups (quantiles) of the data regarding the effect of academic emotions on the achieved grade?"*, the emotions will function as explanatory variables in a regression model. Then the β parameters will be evaluated. Assuming there is a difference between the parameters and thereby the effect of the emotions on the grades, the parameters must significantly differ from the β parameter obtained by the Ordinary Least Squares (OLS) mean regression model. Concluding to the hypothesis that the β_τ parameters obtained by the quantile regression approach are not in the 95% confidence interval of the estimated β by OLS.

The research in gender difference, introduced in the second subquestion, is investigated by the use of an indicator function in the regression model. Thereby it is possible to split the data points into two different classifications of which the quantile regression parameters can be investigated. The hypothesis on the gender difference is therefore, are the $\beta_{Female,\tau}$ parameters significantly different from the $\beta_{Male,\tau}$ parameters. Besides this, are the $\beta_{Female,\tau}$ and $\beta_{Male,\tau}$ parameters significantly different from the β parameter in the union of the distributions.

The research on the long-run effect of the emotions can be investigated by using the significance of the quantile parameters against the OLS parameters in different time periods. Therefore the hypothesis is that the significance of the parameter values obtained by the quantile regression approach are different over time.

4.5. Data Classification

As mentioned, the data used in this research is extracted by dispensing a questionnaire along engineering students, performed by PRIME. Therefore the scale of agreement of the students aligns between "strongly disagree"(1) and "strongly agree"(5). This can be classified as ordinal data. Thus the data is classified in rankings, however the distance in value between these categories is unknown.
The research population consists of first year students in the first semester of cohort 2021/2022 from Civil Engineering, Mechanical Engineering, Applied Earth Sciences and Maritime Engineering following the course *Mathematics 1*. The students were asked to voluntarily take the questionnaires in three different time schedules. One at the start of the mathematics course, one after the first exam and one after the course.
In the first questionnaire at the start of the semester of the course, there were 207 students of which 77 completed the questionnaire. After the exam, there were 121 students of which 40 completed the questionnaire. Finally, after the course of the 105 students 45 responsed the questionnaire.
Of the 60 AEQ questions, 9 questions were related to anger, 6 to boredom, 10 to enjoyment, 6 to hopelessness, 6 to pride, 8 to shame and 15 to anxiety. Besides this, the information on gender of these students is extracted from the questionnaire.
The grade of the first test of the course is used as value for academic performance of the students. This data was available from 114 students.
To visualize the data of the grades attained by the students, figure 4.1 shows the histogram of the density of the attained grade. The histogram of density is created by plotting the frequency of data points within equal intervals of values on a graph. The line is a smoothed version of the histogram used to estimate the probability density function. The program *R* uses the kernel density estimate as non-parametric method to create an estimation of the underlying probability density by the *density()* function.[11] The red dotted line is the mean of the grades.
In Figure 4.1, it is noticeable that the attained grades of the first test are asymmetrically distributed. Furthermore it can be observed that the distribution seems to have two peaks, just before and after the mean. Therefore the distribution can be considered a bimodal distribution.

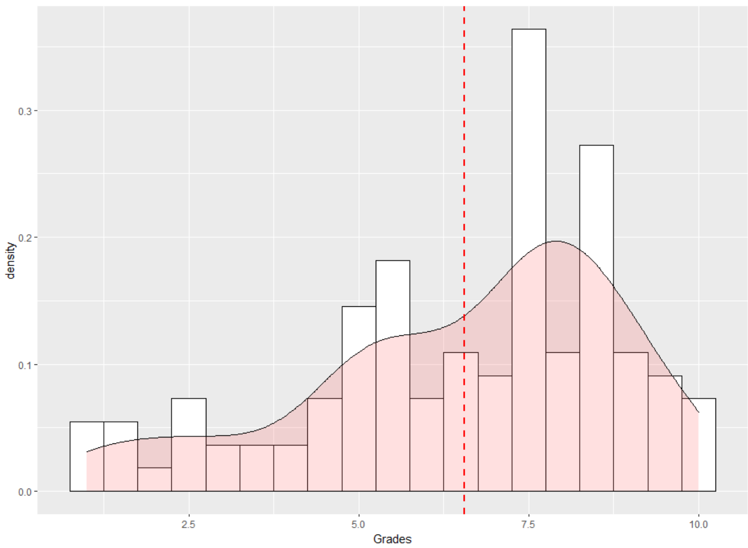

Figure 4.1: Distribution of grades

4.6. Implementation of Quantile Regression

From the estimated probability density function of the grades in Figure 4.1, it is assumed to follow an asymmetric bimodal distribution. By the bimodal characteristics, the data points are not considered to concentrate around the mean. Besides this, the grades are asymmetrically weighted and therefore it is favourable to consider the impact of the covariates on the entire distribution of the grades, instead of the singular conditional mean. As quantile regression allows us to find different conditional effects per fraction of the response variable, it will give a richer characterization of the data compared to conditional mean of the OLS. Therefore quantile regression is a useful method to indicate the relationship of emotions at different percentage groups of the attained grades of the students.

Primarily to detect how the quantiles of the grades behave, the quantiles are plotted against the sorted fractions of the data they correspond with in figure 4.2. So the figure represents the values from the quantile function 2.5 introduced in Chapter 2. The blue dots represent the quantiles for every 10% of the data. The behaviour of the quantiles of the grades can now be examined. There is not a very significant increase in the graph between quantile values, however it can be concluded that there is a sharper increase for the lower fractions of the data. This may suggest a different relationship between the explanatory variables and each conditional quantile of the response variable rather than a link between the expected grade and its mean as ordinary linear regression does.[12]

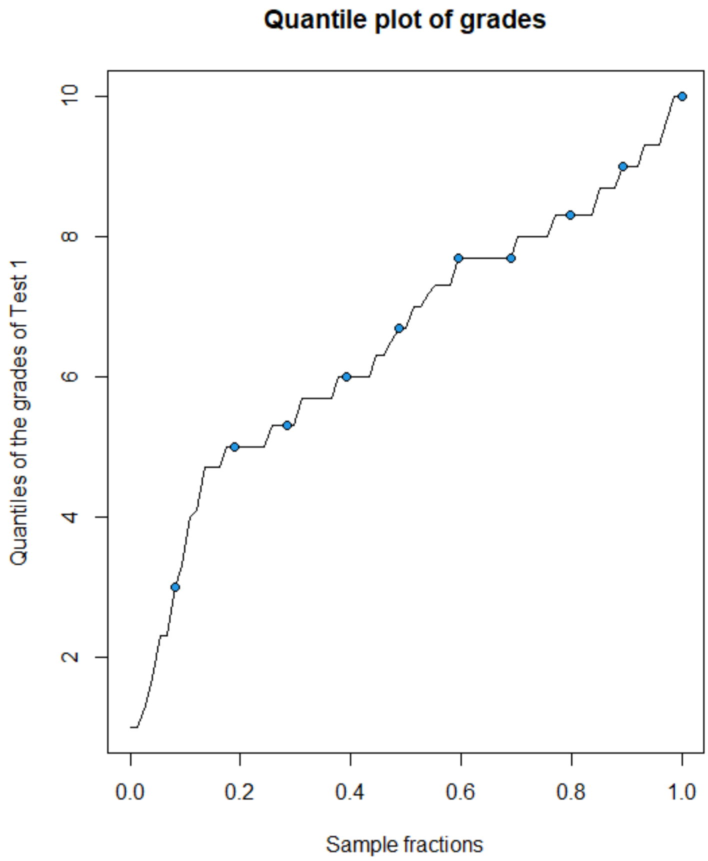

Figure 4.2: Quantiles of Test 1 against fractions of the sample

4.6.1. Univariate Conditional Quantile Regression

To further investigate the relationship between the covariates and the attained grade, the conditional quantile regression function is appropiate. First, the univariate quantile regression is examined, using the *Anxiety* as single covariate. As described, there were multiple questions about *Anxiety* in the questionnaire. Therefore it is suggested to take the mean outcome of all these questions of one individual respondent to be the total *Anxiety value* of this respondent. Therefore the following model of the regression line is examined:

$$G_i = \beta_0 + \beta_1 X_{1i}$$

With X_{1i} equal to the mean anxiety of individual i and G_i defined as the *Attained grades of Test 1* for individual i. Now the goal by using quantile regression is to find the parameters β_0 and β_1 to acquire more knowledge on the effect of *Anxiety*.

To find the parameters for the function, the minimization problem in equation (2.9) is considered. The parameters minimalizing the loss are calculated for several quantiles, finding different parameters per quantile. The intercept is calculated by suggesting the value of the covariate to be equal to zero. The parameter estimators for the univariate *Anxiety* regression are plotted in figure 4.7. In this figure the red line represents the ordinary least squared regression parameter and the red dotted line the 95% interval of this estimation. The parameter solutions minimizing the weighted absolute values are represented by the black dots.

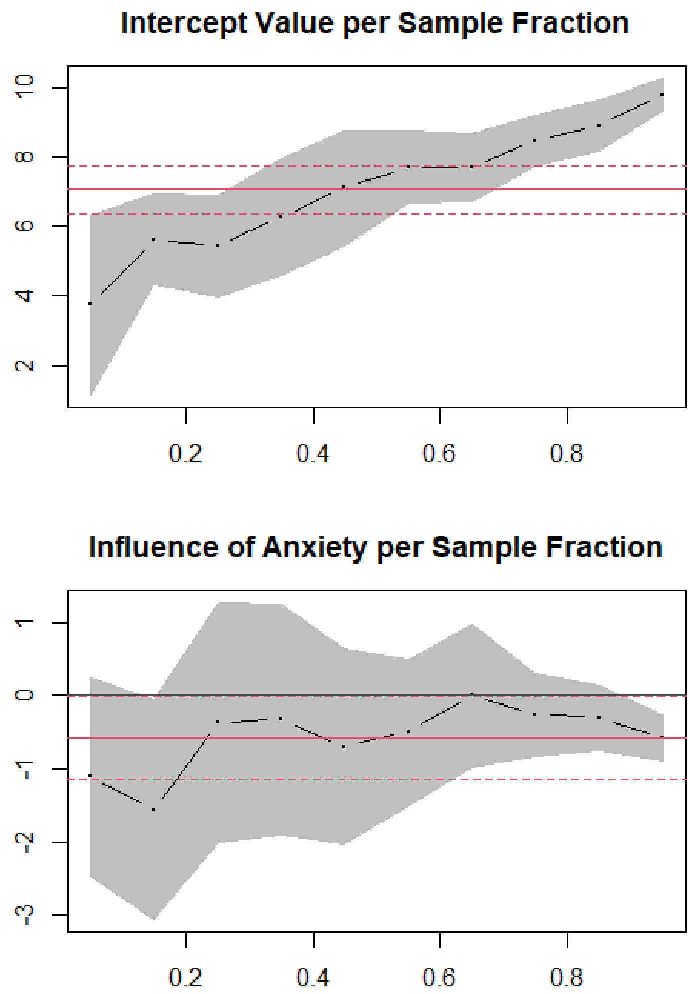

Figure 4.3: Parameter estimators for Anxiety Questionnaire 1

The intuition behind Figure 4.3 is that for a certain quantile, the optimal intercept and β_1 is calculated, such that the conditional quantile is expected to be $G_i = Intercept + \beta_1 * X_{1i}$. Further expansion of the use of this quantile regression in different periods, genders or with multiple emotions can be found in section 4.7.

4.6.2. Multivariate Conditional Quantile Regression

As mentioned in section 4.2 the emotions that were analysed in the questionnaire are anxiety, anger, boredom, enjoyment, hopelessness, pride and shame. Therefore a multivariate quantile regression with seven covariates can be investigated, including the conditional effect of all emotions. Therefore the multivariate regression model including all emotions is:

$$G_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \beta_4 * X_{4i} + \beta_5 * X_{5i} + \beta_6 * X_{6i} + \beta_7 * X_{7i}$$

Although this assumes to give more information on the effect of emotions on the attained grade, it may also be less precise due to multicollinearity. In statistics, multicollinearity is present when one independent variable in a multiple regression model appears to be linearly predicted from the others with a substantial degree of accuracy.[23] Multicollinearity generally does not deduce reliability on the estimation of the model, however it could change the coefficient estimates of the multiple regression erratically in response to small changes in the data. As the main goal of the research is to analyse the effect of emotions on grades rather than estimating the dependent variable, the multivariate conditional quantile regression might be less informative in case of multicollinearity.

4.6.3. Implementation in R

There are several software programs featuring quantile regression commands and procedures, especially statistical programs as Stata, R and SPSS. In this study the R package *quantreg* is used to perform quantile regression as it contains Estimation and Inference methods for conditional quantile functions and is easy to use when familiar with R implementations. The code for the analysis of this data can be found in Appendix A.3.

4.6.4. Data cleaning

Before using the *quantreg* package, for every individual the answered questions of the questionnaire are divided into sets representing the distinct emotions. For every individual in the data set who fully completed the questionnaire, the mean of the answers per set of emotions is calculated. Then as the grades were not collected from the questionnaire, the right grades are connected to the right data of emotions. These are connected by using the anonymised student numbers.

4.6.5. Implementation: Quantreg package in R

As we now have vectors of the same length of the dependent and independent variables, we can use the *quantreg* package in R. The *rq()* function computes and estimates the π-th conditional quantile function of the response variable, given the covariates. It solves the linear programming problem that minimizes the weighted sum of absolute residuals as mentioned in the theory.

The *summary* function returns the estimated intercept and slope coefficient. Intercept represents the estimated quantiles of grades for all covariates being equal to zero. Therefore the intercept is higher for upper quantiles than for lower quantiles of grades.
To gain more information on the parameters, there are several methods for the interference of the coefficients. One of the methods is forming a confidence interval using the bootstrap function *se = boot*, which computes the standard bootstrap errors.

Using the *summary* option together with the wild bootstrap algorithm by *se = boot, bsmethod = "wild"*, the parameters, standard error, t value and probability can be obtained. The standard error of the estimated parameter is the standard deviation of the sampling distribution by the bootstrap method. The standard deviation is the amount of variation in this set of values. Therefore the lower the standard deviation, the closer the values tend to be to the estimated value. In the summary statistics, the confidence interval is then estimated from the distribution of the parameter estimates across the bootstrap samples.

The interpretation of the confidence interval is that *"The 95% confidence interval represents values that are not statistically significantly different from the point estimate at the .05 level"* by Cox D.R. (1974)[5].

4.6.6. Bootstrap function

To find a confidence interval for the parameters of the quantiles, we use the bootstrap function. Bootstrap confidence intervals for β_π are often used as theoretical confidence intervals, but may be hard or impossible to compute analytically. In R we use the *boot.rq* function of the package *quantreg*, which generates standard errors, confidence intervals and tests of hypotheses by bootstrap replicates of the statistic applied to the data. There are five methods available in R for the bootstrap function of which the Wild bootstrap is used in this case.

A Wild bootstrap function, introduced in Chapter 2, is suited when the model exhibits heteroscedasticity and the pattern of heteroskedasticity is unknown. Heteroskedasticity refers to a situation where the variance of the residuals is unequal over a range of measured values. This is true for the present data set as the parameters differ between quantiles, indicating the variance or residuals vary throughout the data.

4.6.7. Dichotomous variables for Gender

To answer the research question, it is desired to investigate the difference in gender. Therefore the data should be divided into gender groups. To this end, a dichotomous variable is formed. A dichotomous variable is a variable that is limited to a fixed amount of possible values. In this case the students were asked for their gender by "Male"(1), "Female"(2), "non-binary"(3), "rather not say""(4), "different, namely"(5). As the students were only male or female, the dichotomous variable can be limited to two values representing the two different gender options. In the linear regression model, this dichotomous variable is represented by an indicator dummy variable. This variable is equal to one if the gender is equal to Female, and to zero otherwise. This leads to the following linear regression model:

$$G_i = \beta_0 + \beta_{gender} D_{1i}$$

$$\mathbf{1}_{D_i} := \begin{cases} 1 & \text{if } i \text{ is Female} \\ 0 & \text{if } i \text{ is Male} \end{cases}.$$

Then the β_{gender} shows the parameter that explains the effect of being a female on the grade attained.

The gender of the respondents was included only in the second questionnaire of the total of three questionnaires. However as the anonymous student number are known, the information about the respondents gender can be used for the other questionnaires as well. However, this resulted in only a few data points of which the emotions and gender were both known.
From the gender extraction we know that 35% of the respondents of the questionnaire were female, 65% were male.
To acquire knowledge on the new regression function including the dummy variable, again *rq()* function is used to form a graph on the intercept and β_{gender} on this variable in Figure 4.11.

Another way of using the dummy variable is to derive more information on the relationship between emotions and the attained grade per gender. This information can be extracted by forming a multiple linear regression model including the dummy variable as well as the independent variables on emotions:

$$G_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_{1i}$$

G_i again being the grade of individual i, β_0 the intercept, $\beta_1 X_{1i}$ the mean anxiety of individual i with its parameter and finally the dummy variable depending on gender of individual i, $\beta_2 D_{1i}$. This results in two different regression lines depending on gender. The dummy variables have the effect of altering the intercept, but the coefficients of the slope do not change. The effect of using the dichotomous variable for the genders can be found in Figure 4.4.

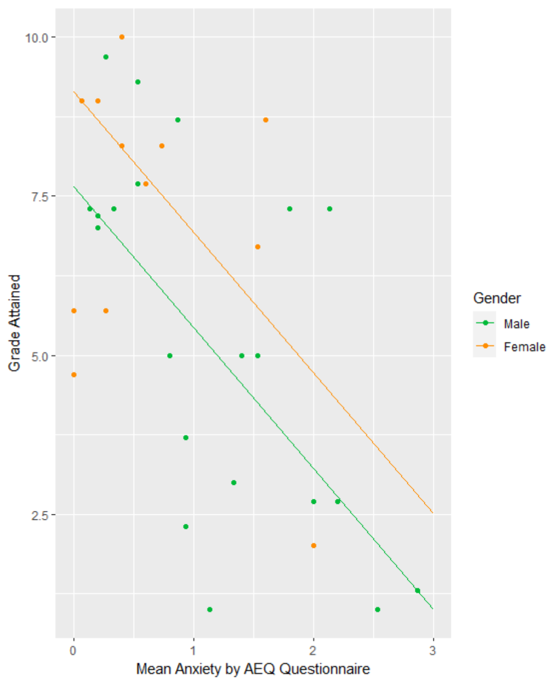

Figure 4.4: Gender difference in affect of anxiety

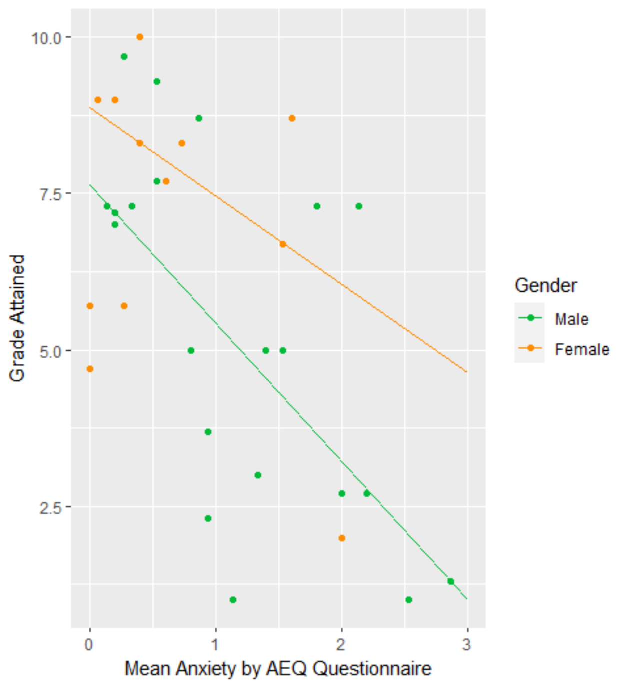

Figure 4.5: Gender difference in affect of anxiety including interaction

If interaction between the gender dummy variable and the anxiety is added in the linear regression, the coefficients of the slopes could change as well. This is shown in Figure 4.5.

$$G_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_{1i} + \beta_3 D_{1i} X_{1i}$$

It can be concluded that by including the interaction of the dichotomous variable with the other independent variable, the quantile regression model fits the data better.

4.7. Results

To answer the research questions in section 4.3, the data is investigated in several manners. First the effect of anxiety on the students' attained grades is researched, as anxiety is a common stress factor of which the understanding of the influence on education could be important. The quantile regression indeed shows different β parameters for different quantiles.
After looking at one specific emotion, a multivariate model of multiple emotions is formed to find results on the influence of all studied emotions by the AEQ on the attained grades of the students. As shown earlier, the gender might have an influence on the grades of the students. Therefore also the gender is further investigated to answer the research question, as the gender distributions act very differently. From the solutions of the gender difference, a simpson's paradox may be of topic.
Finally the change over time during the semester of the students following the mathematics course will be elaborated.

4.7.1. Anxiety

To investigate the relationship between anxiety and attained grades of the students, for every individual the grades are plotted against the mean of all anxiety questions in Figure 4.6. The Ordinary Least Squared (OLS) regression line and the Quantile regression line for the median are included, to see whether there is a relation between the two variables.
In Figure 4.6 it is noticeable that for the OLS regression as well as for the quantile regression, the coefficient of the regression line is negative. Implying that a higher anxiety causes a lower attained grade. It can be concluded that a regression line that is more robust to outliers, the median quantile regression line, has a more negative regression line. The coefficient of the OLS line is -0.58, while the coefficient of the median quantile regression is -1.0. Causing the estimation of the effect of the anxiety to be of bigger influence. Besides

this the p-value of the median quantile regression is 0.081, where this is 0.094 for the OLS regression. Although both are not considered significant for the $\alpha = 0.05$, the median quantile regression gives a smaller p-value.

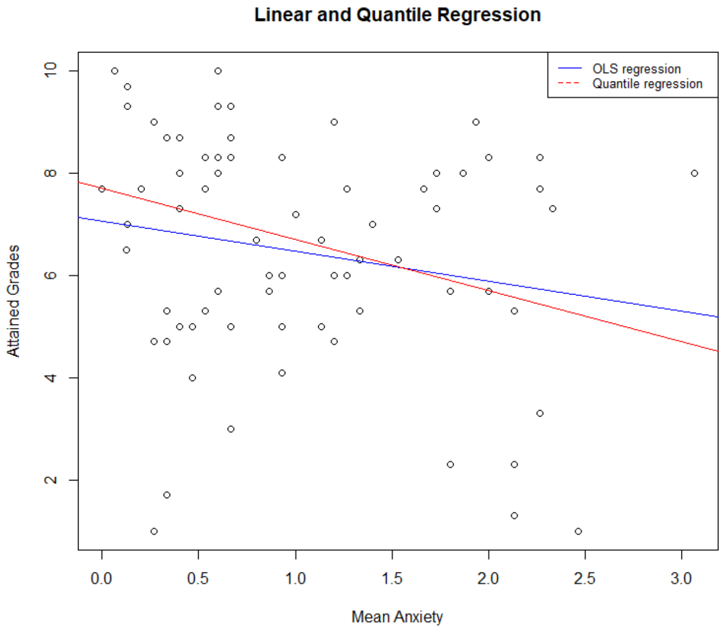

Figure 4.6: Ordinary Least Squared and Quantile Regression on the relationship between anxiety and grades

To see whether the effect of the anxiety differs over the fractions of data, we investigate the parameters of anxiety by quantile regression.
In Figure 4.7 the red line is the OLS estimation. The dotted red line shows the 95% confidence interval of this OLS regression coefficient. The horizontal axis shows the quantiles of the response variable and the vertical axis are the coefficient magnitudes of the explanatory variable. The black dotted line shows the coefficients of the quanatile regression and the gray boundaries represent the 95% interval of this quantile regression coefficient. Therefore if 0 is not included in the 95% interval, the coefficient is significantly different from 0 as its p-value is below $\alpha = 0.05$. The results and p-values of the OLS regression and quantile regression can be found in Table 4.7.1 (* showing significance). From Figure 4.7, it can be seen that the influence of Anxiety is expected to differ between the different quantiles of the research population. As the dotted line represents the 95% confidence interval of the OLS estimation, it can be concluded that the influence of Anxiety for the 0.15 quantile is significantly different from the OLS estimation.

Coefficients for different quantiles			
Total Grade	Intercept	Anxiety coefficient	P-value
OLS regression	7.0560	-0.5833	0.0938
$\tau = 0.05$	3.74074	-1.11111	0.08817
$\tau = 0.15$	5.62308	-1.55769	0.03194*
$\tau = 0.25$	5.42500	-0.375	0.65384
$\tau = 0.35$	6.27857	-0.32143	0.70071
$\tau = 0.45$	7.09286	-0.69643	0.28937
$\tau = 0.55$	7.70000	-0.50000	0.39349
$\tau = 0.65$	7.70000	0.0000	1.00000
$\tau = 0.75$	8.45882	-0.26471	0.51713
$\tau = 0.85$	8.90000	-0.30000	0.40616
$\tau = 0.95$	9.77727	-0.57955	0.00425 *

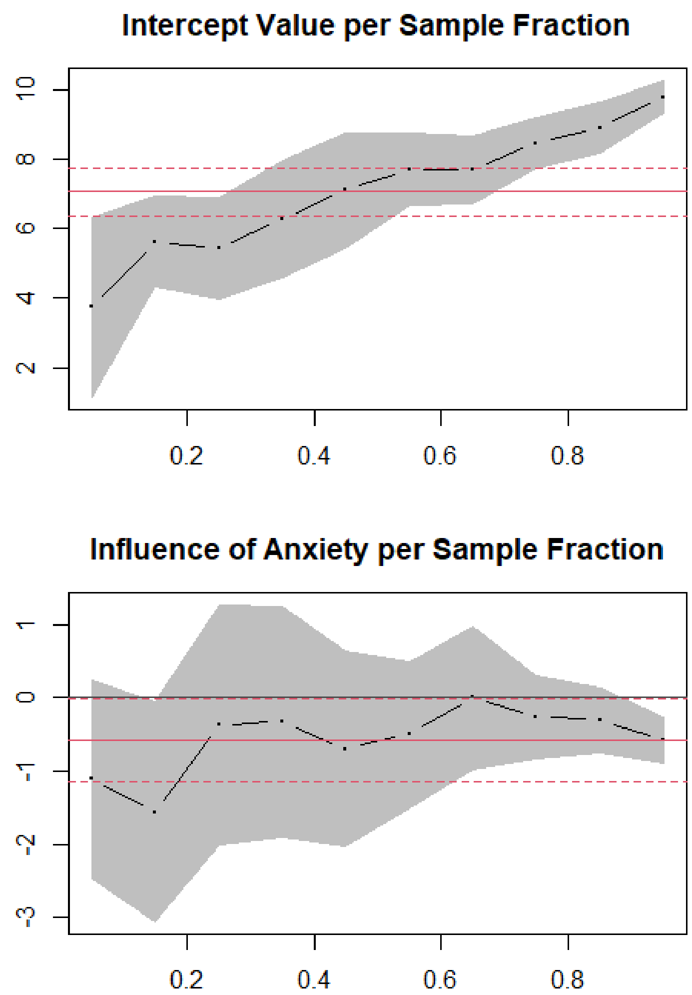

Figure 4.7: Parameter estimators for Anxiety Questionnaire 1

4.7.2. Multiple Emotions

To research the regression for all the emotions, the multivariate conditional quantile regression formed in Chapter 4.6.2 is implemented. The parameters are similarly extracted using the *quantreg* package for conditional quantile regression. The behaviour of the parameter estimations when all emotions are included are shown in Figure 4.8

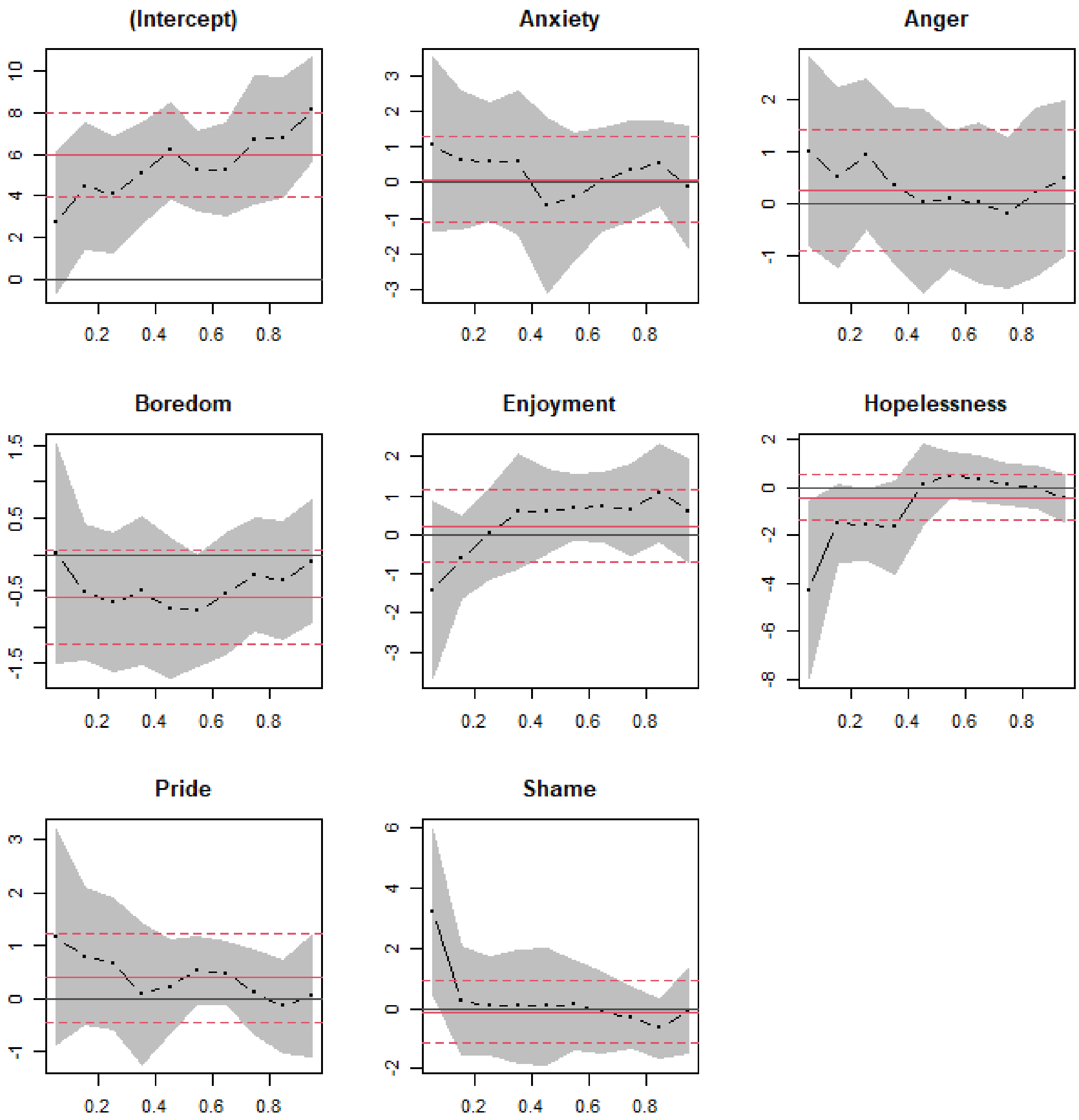

Figure 4.8: Parameters of all emotions

Figure 4.8 can be interpreted as the coefficient Figure in previous subsection. The confidence intervals seem relatively large, suggesting the parameters vary a lot. Significantly, it can be concluded from the figure that the β_1, which is the parameter representing anxiety, is not negatively related to the grades. Earlier in the univariate regression model of anxiety, it did conclude a negative relation. This difference in influence can be due to multicollinearity, as explained in chapter 4.6.2. Therefore the correlation between the explanatory values can be found in Figure 4.9.

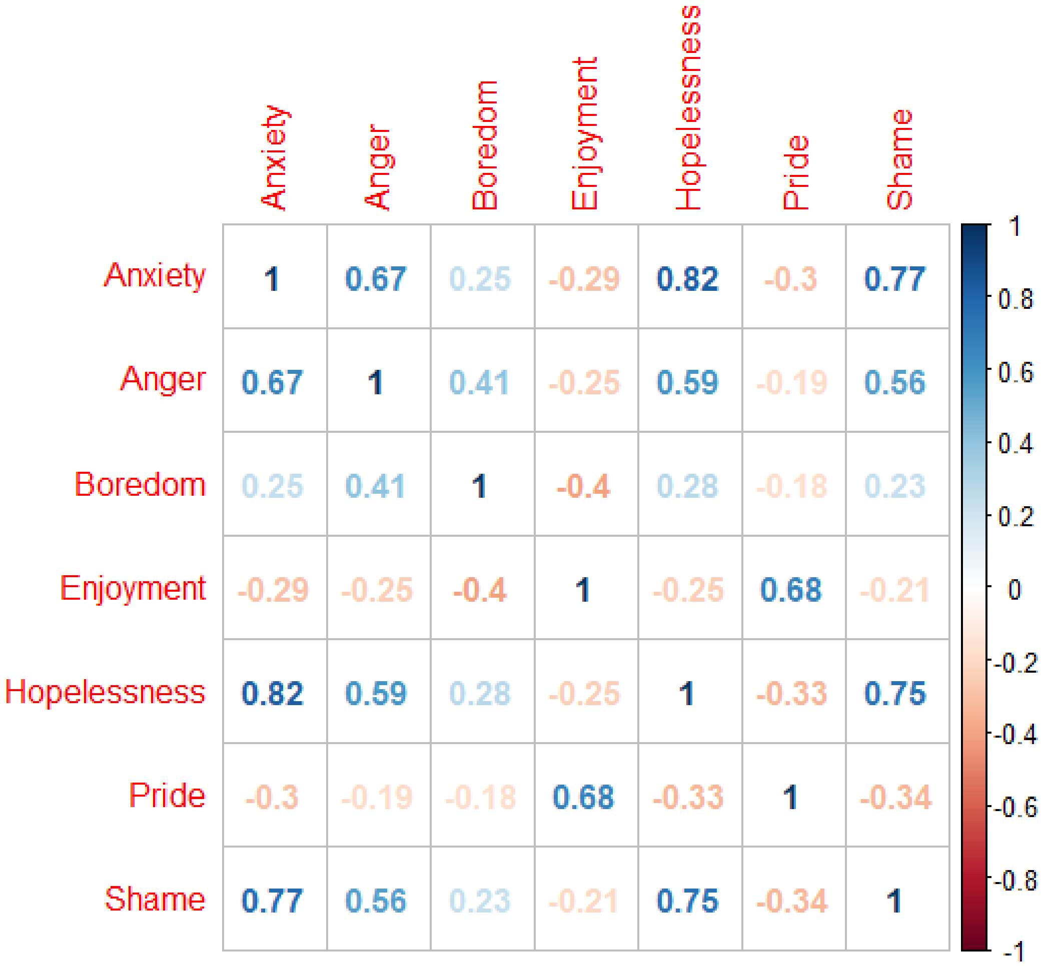

Figure 4.9: Correlation between emotions

As can be seen from the correlation table, the correlation between anxiety and anger, hopelessness and shame are very high. Therefore this multicollinearity can change the coefficient estimates of the anxiety parameter erratically in response to small changes in the model or the data. However, as multicollinearity assumes the results are not valid for individual predictors, another view of interpreting the change is that it may be a result of an estimation of the slope parameter with omitted variable bias in the univariate model. Meaning that the variation in the response variable that is due to another cause is inappropriately linked to the independent variable.
However in this multivariate model, the coefficients of shame and hopelessness of the 0.05 fraction of the response variable are significantly different as the 0 is not included in the 95% interval of the parameters. Besides this, they are also outside the OLS 95% interval.

4.7.3. Gender

The following section includes investigation on the difference of the effect of emotions on attained grades between genders. For this research, the dummy variable introduced in Chapter 2 is used to indicate the genders. To get an intuition on the behaviour of the attained grades of the different genders, the smoothed *density()* function is again used to estimate the probability density, now on gender divided sets.

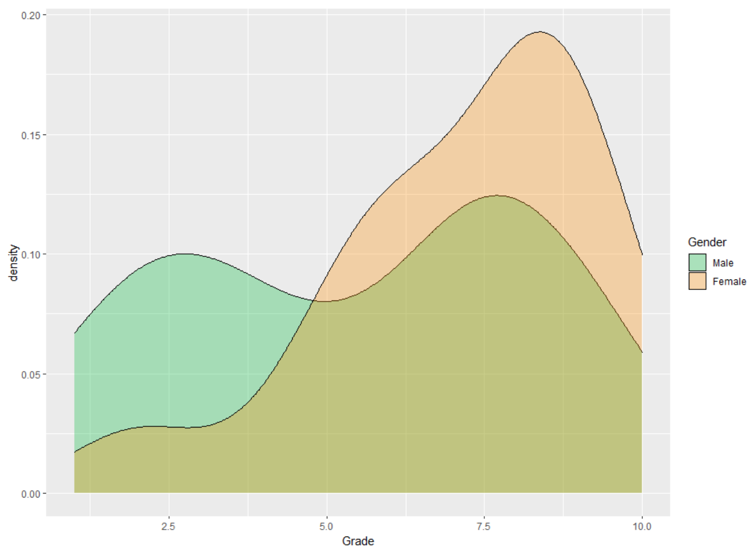

Figure 4.10: Difference Gender

From the Figure 4.10 it can be concluded that the probability density functions act disparate. Therefore further investigation in the effect of gender may be useful for the conclusion.

Using the indicator dummy variable, the result of the Gender effect on the attained grade is shown in Figure 4.11. As assumed from the distribution, the β_{gender} has a positive value, assuming being a female has a positive effect on the attained grade. However, only information on 40 students of whom both grade and gender were known are included of which only 14 were female.

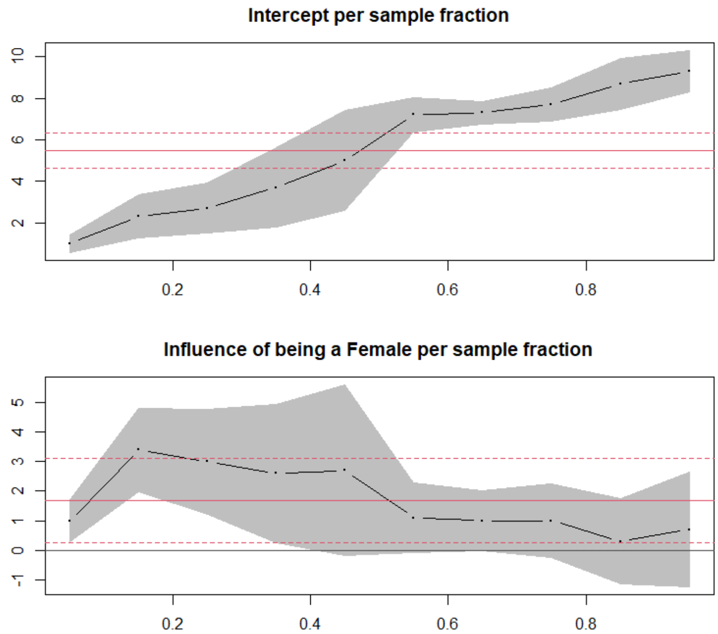

Figure 4.11: Intercept and Beta on Regression using Dummy variable for Gender

To see the effect of anxiety within the different sexes, also a multivariate model is investigated including anxiety and the dummy variable of the gender. In 4.12 the multivariate model $G_i = \beta_0 + \beta_1 X_{1i} + \beta_{gender} D_i$ is shown. As the effect of anxiety within the gender is also of importance, the model including the intersection of both, $G_i = \beta_0 + \beta_1 X_{1i} + \beta_{gender} D_i$, is shown in Figure 4.13.

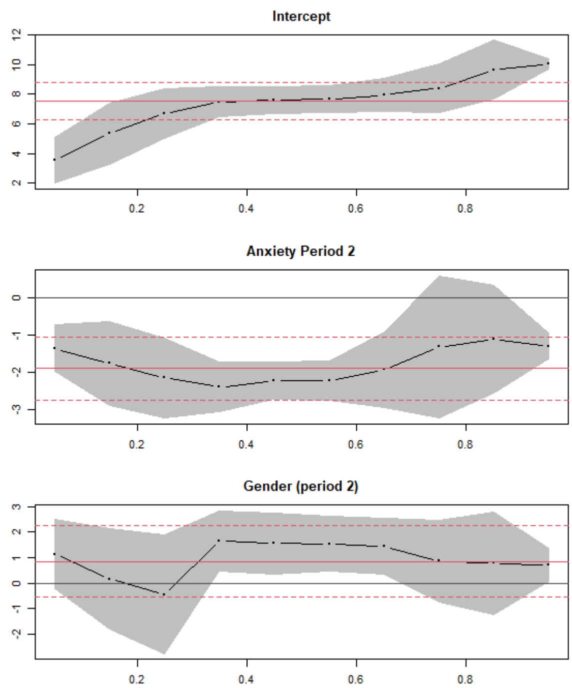

Figure 4.12: Anxiety plus Gender

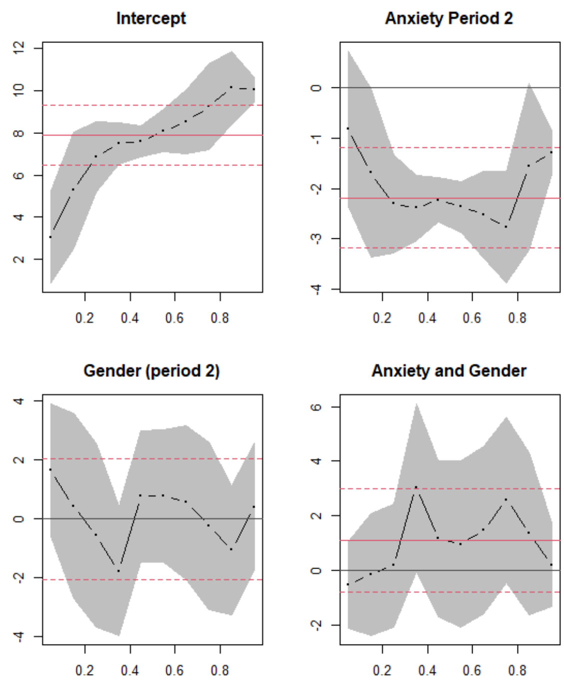

Figure 4.13: Anxiety and Gender interaction

The result of this investigation is quite significant. In Figure 4.12, the anxiety has a negative effect on the attained grade and the gender has an overall positive effect. This is in line with the conclusion with the seperate univariate models. However, in Figure 4.15, the anxiety has a positive effect when the indicator function is

equal to 1. Therefore it can be concluded that female students experience a less negative or sometimes even positive effect of anxiety on their grades.
To get a better intuition of this effect, the coefficient lines of both genders for different quantiles are plotted in Figure 4.14-4.17.

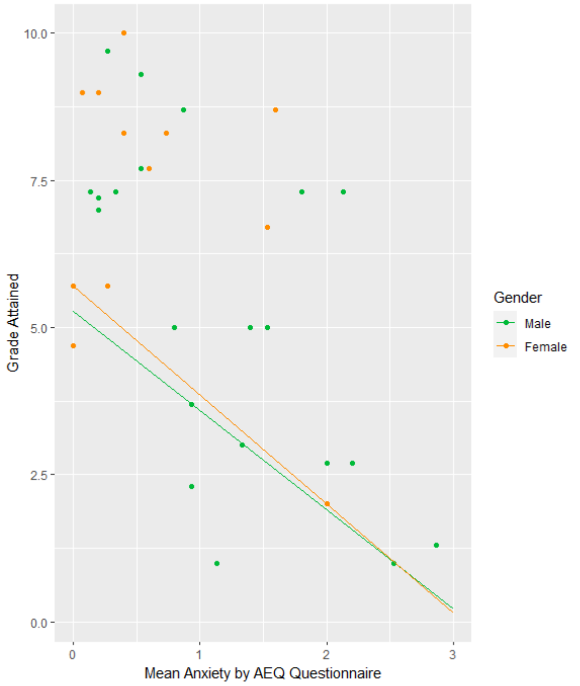

Figure 4.14: Regression line $\pi = 0.15$

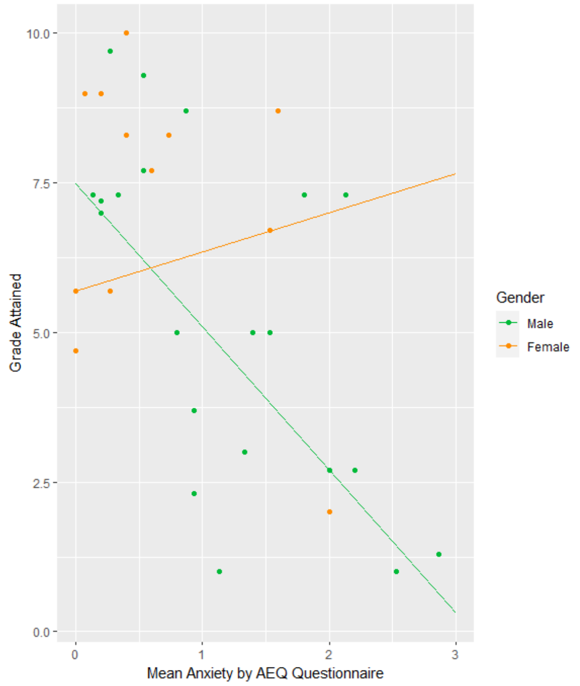

Figure 4.15: Regression line $\pi = 0.35$

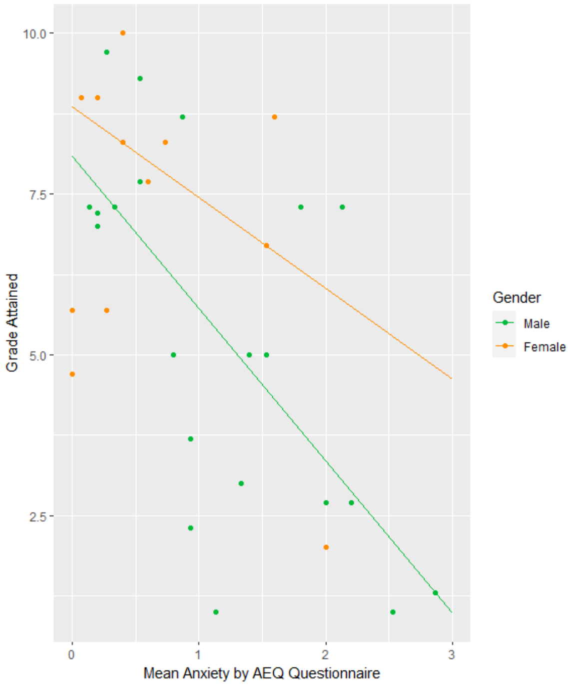

Figure 4.16: Regression line $\pi = 0.55$

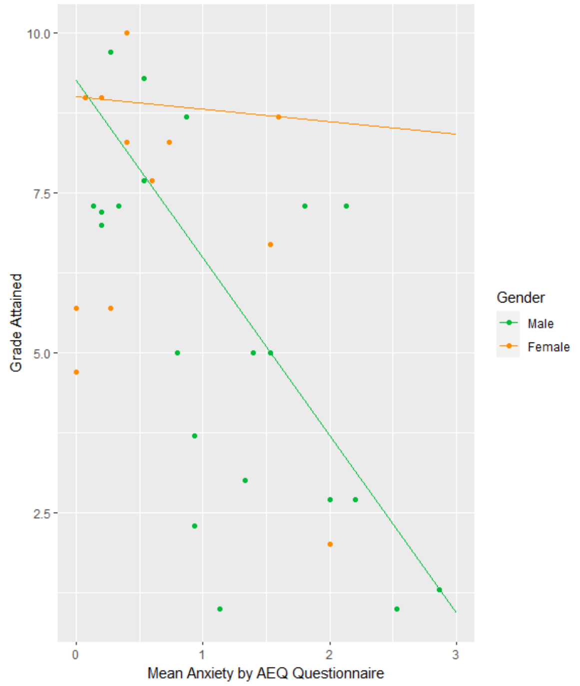

Figure 4.17: Regression line $\pi = 0.75$

This occurance is an example of the Simpson's paradox.

Definition 11 (Simpson's Paradox) *The Simpson's paradox is a phenomenon in probability and statistics in which a trend appears in several groups of data but disappears or reverses when the groups are combined.*

Therefore the trend that is explained by researching the genders seperately is different from when the groups are united. This as the effect of the emotions have different levels for the different genders. Figure 4.18 gives an illustration to explain the simpson's paradox. The dotted line gives a trend when the data is united, which are very different when the groups are seperate.

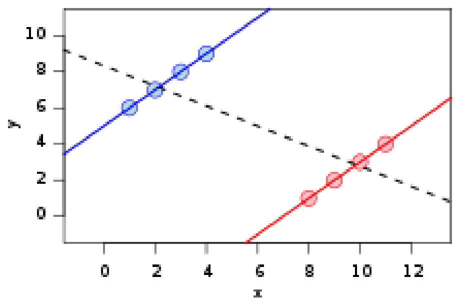

Figure 4.18: Simpson's paradox example

To handle this paradox, causal relations and con-founders must be well managed in the statistical modelling. "Mathematician Jordan Ellenberg argues that Simpson's paradox is misnamed as "there's no constriction involved, just two different ways to think about the same data" and suggests that its lesson "isn't really to tell us which viewpoint to take but to insist that we keep both the parts and the whole in mind at once."[13]"

4.7.4. Emotions over time

The questionnaires were distributed at three different time points. One at the start of the course, one after the first test and one after te course. Due to less responses in the last questionnaires, the multivariate models were responding very radically to small changes in the data. Therefore the univariate model in the analysis of the emotions over time may be more valuable as multicollinearity does not appear.
The effect of anxiety on different quantiles just after the exam are plotted in Figure 4.19. The parameters of the quantile regression line on the data of the questionnaire taken after the course, can be found in Figure 4.20.

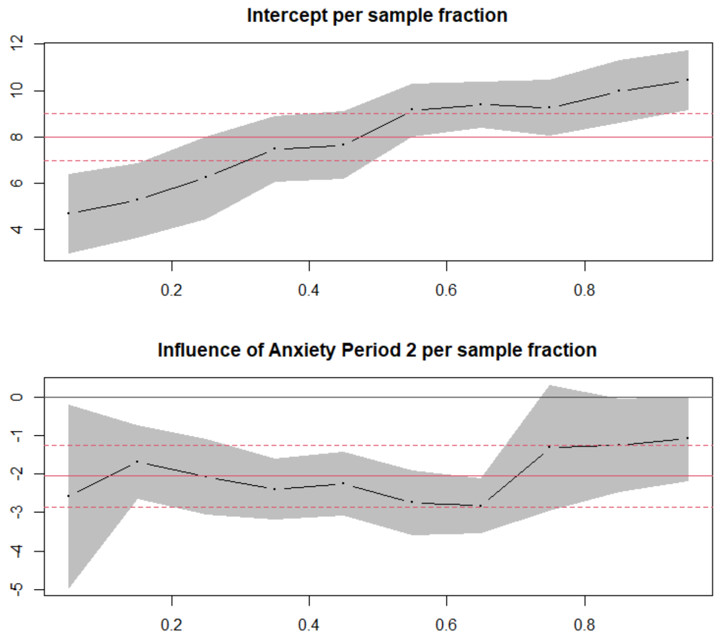

Figure 4.19: Influence Anxiety Period 2

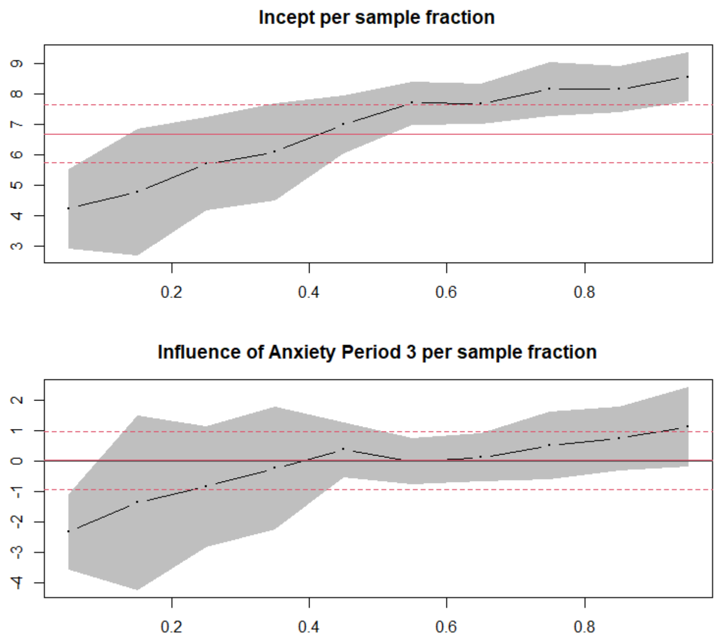

Figure 4.20: Influence Anxiety Period 3

Compared to the effect of anxiety in period one, which is shown in Figure 4.7 in Section 4.7.1, the results seem to be less negative. The higher quantiles of the data set are even expected to experience a positive effect of anxiety on their grades.

4.7.5. Bayesian quantile regression for ordinal longitudinal PRIME data

Using Bayesian quantile regression for ordinal longitudinal data is a simulation that makes a probability distribution for possible ordinal outcomes. Therefore in the simulation using the data of PRIME, the emotions are considered to be the response variable and the latent variable is as follows:

$$l_{ij} = x_{1ij}\beta_1 + x_{2ij}\beta_2 + \varepsilon_{ij},$$

Now the independent vectors are for instance the gender or attained grade. Then the emotions are sampled according by the latent variable. The τ-th quantile regression is then considered as:

$$Q_{Y_i|X_i}(\tau) = \alpha_i + \beta_0 + \beta_1 xGender_{ij} + \beta_2 xGrade_{ij}$$

As explained, for Bayesian quantile regression, prior knowledge can be used to have a better prediction. Therefore the data of the questionnaire is analysed, the histogram of the anxiety emotions can be found in Figure 4.21.

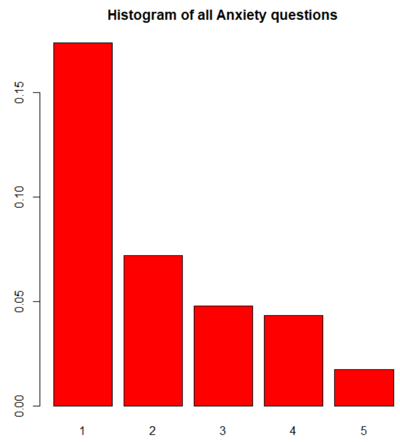

Figure 4.21: Histogram of all anxiety questions

By the figure the prior of an exponential function can be suggested to analyse the emotions as response variable in the Bayesian approach.

5 Conclusion

In this research, several approaches of quantile regression analysis for estimating the effect of emotions on attained grades of engineering students are considered. The general quantile regression approach as well as the Bayesian quantile regression approach is elaborated. To illustrate the use of the information that can be attained by such models, it is applied to data of PRIME.

The quantile regression approach is considered to be robust to outliers and asymmetrical distributions and may discover relationships between variables for abnormal connectivity as it predicts the conditional distribution for different quantiles. The model predicts the parameters by minimizing the loss function, which is a weighted LAV model. The outcome of the model are parameter etimates per quantile that show the relationship between the explanatory variable connected to the parameter value and the response value. The parameters can be resampled with the Wild bootstrap method, such that a 95% interval of the estimates is predicted and significance can be tested. This resampling method is based on the response variable and keeps the indipendent values on their initial value, which is useful as it is a conditional distribution.

The Bayesian approach allows the quantile regression to take into account prior knowledge by first stating a prior distribution. Then as the minimization of the loss function is equivalent to the maximization of the likelihood of independently distributed Asymmetric Laplace densities, the likelihood is calculated by parametrization of the distributions. Furthermore several MCMC methods can be used to form a reliable posterior distribution by the likelihood and prior distribution, of which the Gibbs sampling method is a useful method for ordinal longitudinal data. The posterior mean and variance of the regression coefficients and parameters can then be used to illustrate the relation between the regression variables.

To research the effect of academic emotions on attained grades of engineering students, the method of quantile regression modelling was applied to the data of PRIME. The data was difficult to process and all the information was only available of a few engineering students. However it was possible to see some effects of the emotions on attained grades. Although the multivariate model including all emotions is not considered reliable because of the highly correlated explanatory variables, the quantile regression on the univariate model showed some significant effects. As the parameter value of lower quantiles of the response variable was significantly lower, it can be concluded that students attaining a lower grade experience a more negative effect of anxiety.
Introducing a dummy variable for gender showed another conclusion. The influence of anxiety was assumed to have an overall more negative effect, however the combination of being anxious and female was surprisingly positive for some quantiles. In that case anxiety is assumed to have a positive effect on the attained grade by being more prepared or focused for their test due to anxiety. This was not considered by the OLS regression and assumes the Simpson's Paradox as this effect was not found when the gender groups were combined. Therefore the use of quantile regression was indeed very useful to gain more information and to reconsider conclusions that would have been drawn from the normal OLS regression approach.

Another way of results interpretation might be that females are anxious although their grades are higher. This as the overall distribution of the female scores assumed higher scores than for male students. In that case it is concluded that anxiety seems to exist for any grade. Therefore a new hypothesis about quantiles of emotions might be interesting to investigate, as they may act differently. Therefore the Bayesian quantile regression approach for ordinal longitudinal data was also introduced. In this case, the emotions (ordinal data) are considered the response variable and the gender and scores are the explanatory variables. Therefore this might be interesting to gain more information on how the quantiles of the emotions act. By the conclusion of the Simpson's paradox, it might be useful to approach the data in different gender groups for modelling Bayesian quantile regression. Besides this, the anxiety emotions are considered to act exponentially. Therefore an exponential prior distribution can be used to analyse the emotions.

An overall conclusion is that the quantile regression approach gave new insights and significancy results that would not have been shown by OLS regression. Also seperating genders in modelling quantile regression assumed new conclusions on the effect of emotions, namely positive effects of anxiety. As Bayesian statistics is a reliable way to estimate a posterior distribution by using prior information, likelihood estimates and the MCMC method. This is a good analysis method to further conduct on the quantile regression approach on the effect of academic emotions on attained grades of engineering students.

6
Discussion

During the research there were some complications and limitations. Recommendations on future research are included in this chapter, based on the challenges and limitations of this research.

6.1. Bayesian quantile regression

Although the Bayesian quantile regression for ordinal longitudinal data was investigated and analysed in detail, the method was not applied on the actual PRIME data. Due to the many different insights the quantile regression method offers, the analysis on the data by the regular quantile regression method took a lot of time. Researching it for different genders, time periods and multivariate models was very informative but also time consuming. Therefore the Bayesian quantile regression approach is not applied in R using the data of PRIME, although by the literature of the Bayesian quantile regression it is assumed to be very reliable and informative. Therefore it may be very interesting for further research to construct the Bayesian quantile regression approach and compare it with the general quantile regression.
Besides this, the Bayesian quantile regression for ordinal longitudinal data is explained. In that case the academic emotions would be considered as the response variable. Therefore further research may see the effects from another view. As the emotions may be effected by academic achievements and gender, instead of looking at the academic emotions as an explanatory variable. Also, Bayesian quantile regression may be better for further research as previous information and knowledge can be used for computing the posterior distribution.

6.2. Multicollinearity in multivariate model

As the analysis on the explanatory variables showed a high correlation between the explanatory variables, the multivariate model can be concluded to be unreliable. However a multivariate model without multicollinearity may be more informative then a univariate model, as the variation in the univariate model that is due to another cause can be inappropriately linked to the independent variable. Therefore for future research, a new multivariate model without multicollinearity is suggested, as multicollinearity can be avoided by linearly combining predictor variables.

6.3. Clusters

In the quantile regression analysis, many clusters are combined regarding studies, age, gender and prior education. Clustering the gender gave us new insights on the relationship between the variables investigated in the quantile regression. Although clustering is difficult when there is not a lot of data, for further research it may be interesting to use clustering for studies, age and prior education as well. These influential factors may influence groups which are assumed to be independent in this research.

6.4. Data

As mentioned, it is difficult to cluster when not a lot of data is available, as the effect of influential factors cannot be shown as the groups become very small. The data of the questionnaires was not easy to extract and connect to eachother. Therefore for further investigation, it is advised to reconsider the data ordening. It is useful to extract more data for further investigation in Bayesian quantile regression, as it allows more ordinal longitudinal data points. The data on gender was only available in the second questionnaire, which was the least answered questionnaire. When more data is available, the conclusions might also be more reliable.

A
Code

A.1. Distribution of grades

```
#Data van eerste vragenlijst
vragenlijst_1_cijfers
vragenlijst_1_cijfers[1,1]

#Data overzetten in Matrix
vragenlijst1 <- str_split_fixed(vragenlijst_1_cijfers[1:207,1],",",137)
# rij is de leerling, kolom de vragen
vragenlijst1

vragenlijst2 <- str_split_fixed(vragenlijst_2_cijfers[1:121,1],",", 141)
vragenlijst2

vragenlijst3<- str_split_fixed(Vragenlijst_3_cijfers[1:105,1], ",", 138)
vragenlijst3

AnxietyKolommen <- c(56,57,58,61,75,77,80,84,92,95,97,99,101,105,109)
TotalAnxiety = data.frame(NA)

#histogrammen
anxietyVraag2 = vragenlijst1[,56]
for (i in AnxietyKolommen){
  strQuestion <- as.character(i-54)
  strMain <- paste("Histogram question ", strQuestion)
  barplot(table(as.numeric(vragenlijst1[,i]))/length(as.numeric(vragenlijst1[,i])), main = strM
  TotalAnxiety<-cbind(TotalAnxiety, as.numeric(vragenlijst1[,i]))
}
return(TotalAnxiety)

numTotalAnxiety <- stack(TotalAnxiety)
barplot(table(as.numeric(numTotalAnxiety[,1]))/length(numTotalAnxiety[,1]), main = "Histogram o

median(as.numeric(numTotalAnxiety[,1]), na.rm = TRUE)
median(as.numeric(vragenlijst1[,101]), na.rm = TRUE)
TotalAnxiety[21,]
#conclusie: over het algemeen zijn de anxiety vragen ongeveer exponentieel verdeeld


#cijfers van leerlingen
Anonieme_cijfers
DataCijfers <- str_split_fixed(Anonieme_cijfers[1:114,1], ",",12)
DataCijfers
```

```
#Histogram Cijfers om distribution te bepalen
# Histogram overlaid with kernel density curve
ggplot(as.data.frame(DataCijfers[,9]), aes(x= as.numeric(DataCijfers[,9])))+
  geom_histogram(aes(y=..density..),      # Histogram with density instead of count on y-a
                 binwidth=.5,
                 colour="black", fill="white") +
  geom_density(alpha=.2, fill="#FF6666") +
  geom_vline(aes(xintercept=mean(as.numeric(DataCijfers[,9]), na.rm=T)),   # Ignore NA val
             color="red", linetype="dashed", size=1) +
  xlab("Grades")
```

A.2. Quantiles of grades

```
#STEP 1 QUANTILES OF GRADES
#to see how the variables quantiles behave
cijfers <- as.numeric(MatrixAnxietyGrades[,2][!is.na(MatrixAnxietyGrades[,2])])
n <- length(cijfers)
plot((1:n-1)/(n-1), sort(cijfers), type = "l", ann = FALSE)
title("Quantile␣plot␣of␣grades", xlab = "Sample␣fractions", ylab = "Quantiles␣of␣the␣grade
points(((1:n-1)/(n-1))[seq.int(0, n, n/10)], sort(cijfers)[seq.int(0, n, n/10)], pch = 21,
```

A.3. Quantile Reg: Anxiety against grades

```
################################################
FORMS MEAN OF ANXIETY QUESTIONS PER STUDENT
################################################
AnxietyKolommen <- c(56,57,58,61,75,77,80,84,92,95,97,99,101,105,109)
#Calculates the mean anxiety on all questions of a particular student
meananxiety = data.frame(NA)
for (i in 1:207){
  meananxiety[i] = mean(as.numeric(vragenlijst1[i, AnxietyKolommen]), na.rm = TRUE)-1
}

####################################################
#Forms data for plot of anxiety against grades
####################################################
nmmr=0
MatrixAnxietyGrades = matrix(NA, nrow = 80, ncol = 4)
for (k in 1:114){
  for (l in 1:207){
    #if the student numbers are identical
    if (DataCijfers[k,2] == vragenlijst1[l,2]){
      #they are added to the new dataset
      nmmr <- nmmr +1
      if (is.na(as.numeric(DataCijfers[k,9])) == FALSE){
        if (is.na(as.numeric(meananxiety[l])) == FALSE){
          MatrixAnxietyGrades[nmmr, 1] = DataCijfers[k,2] #student nr in col 1
          MatrixAnxietyGrades[nmmr, 2] = DataCijfers[k,9] #grade T1 in col 2
          MatrixAnxietyGrades[nmmr,3] = as.character(meananxiety[l]) #mean anxiety col 3
        }
      }
      }
  }
}

x <-MatrixAnxietyGrades[,3] #meananxiety
y <-MatrixAnxietyGrades[,2] #grades
```

```
x <- as.numeric(x[!is.na(x)])
y<- as.numeric(y[!is.na(y)])

summary(lm(y~x))
summary(rq(y~x), se="boot", bsmethod = "wild")

plot(y ~ x, main = "Linear␣and␣Quantile␣Regression",
     xlab = "Mean␣Anxiety",
     ylab = "Attained␣Grades")
abline(lm(y ~ x), col = "blue", xpd = FALSE) #linear regression
abline(rq(y ~ x), col = "red", xpd = FALSE) #quantile regression
legend("topright", legend=c("OLS␣regression", "Quantile␣regression"),
       col=c("blue", "red"), lty=1:2, cex=0.8)

##########################################
#QUANTILE REGRESSION ANXIETY 1
##########################################
QR_GRADES <- rq(y~x, data = as.data.frame(MatrixAnxietyGrades),
tau = seq(0.05,0.95,by=0.1))
QR.1 <- summary(QR_GRADES, se = "boot", bsmethod = "wild")
plot(QR.1, main = c("Intercept␣Value␣per␣Sample␣Fraction",
                    "Influence␣of␣Anxiety␣per␣Sample␣Fraction"),
     xlab = "Sample␣Fraction",
     ylab = "Beta")
```

A.4. Quantile Reg:Gender analysis

```
###################################################
#Forms data for plot of sex against grades
###################################################
nmmr=0
MatrixSexGrades = matrix(NA, nrow = 80, ncol = 4)
for (k in 1:114){
  for (l in 1:121){
    #if student numbers are equal in both data sets
    if (DataCijfers[k,2] == vragenlijst2[l,2]){
      #they are added to the new dataset
      nmmr <- nmmr +1
      if (is.na(as.numeric(DataCijfers[k,9])) == FALSE){
        if (is.na(as.numeric(vragenlijst2[l,6])) == FALSE){
          MatrixSexGrades[nmmr, 1] = DataCijfers[k,2] #student number in col 1
          MatrixSexGrades[nmmr, 2] = DataCijfers[k,9] #grades attained in col 2
          MatrixSexGrades[nmmr,3] = as.character(vragenlijst2[l,6]) #sex in col 3
        }
      }
    }
  }
}

Sex <- as.factor(MatrixSexGrades[,3][!is.na(MatrixSexGrades[,3])])
Gradesex<- as.numeric(MatrixSexGrades[,2][!is.na(MatrixSexGrades[,2])])

QR_SEX <- rq(Gradesex~Sex, data = as.data.frame(MatrixSexGrades), tau = seq(0.05,0.95,by=0.10))
QR.3 <- summary(QR_SEX, se = "boot", bsmethod = "wild")
plot(QR.3, main = c("Intercept␣per␣sample␣fraction", "Influence␣of␣being␣a␣Female␣per␣sample␣fr

#################################################################
Gender difference Grade against Anxiety
#################################################################
    GenderModel <- rq(GradeEmotions2 ~ Anxiety22 + Gender + (Anxiety22*Gender), data = as.data.
```

```
mydata <- tibble(Anxiety22, Gender, GradeEmotions2)
lines_gender <- tribble(~Anxiety22, ~Gender,
                        0,1,
                        3,1,
                        0,2,
                        3,2) %>% mutate(
                          Gender=factor(Gender)) %>% mutate(
                            GradeEmotions2=predict.rq(object = GenderModel,
                                             newdata = .))

ggplot(data= mydata,
       mapping=aes(x=Anxiety22,
                   y= GradeEmotions2,
                   color=Gender)) + scale_color_manual(values = c("#00BA38","#FF8D00"),
                                                       name = "Gender",
                                                       breaks = c("1", "2"),
                                                       labels = c("Male", "Female"))+
  geom_point() +
  geom_line(data= lines_gender %>% filter(Gender==1), na.rm = TRUE) +
  geom_line(data= lines_gender %>% filter(Gender==2), na.rm = TRUE) +
  xlab("Mean␣Anxiety␣by␣AEQ␣Questionnaire") +
  ylab("Grade␣Attained") +  scale_shape_discrete(name = "Gender",
                                                 breaks = c("1", "2"),
                                                 labels = c("Male", "Female"))
```

A.5. Quantile Reg:Multiple emotions

```
######################################
#QUANTILE REGRESSION ALL EMOTIONS
######################################
AngerKolommen <- c(63, 67,  69,  71,  78,  83,  88, 102, 107)
BoredomKolommen <-c(6,11,14,20,25,31)+54
EnjoymentKolommen <- c(1,5,8,12,22,28,35,39,46,54)+54
HopelessnessKolommen<-c(40,42,44,49,52,56)+54
PrideKolommen<-c(18,19,32,36,58,59)+54
ShameKolommen<-c(10,16,27,33,37,50,57,60)+54

AngerKolommen2 <- c(63, 67,  69,  71,  78,  83,  88, 102, 107)+ 4
BoredomKolommen2 <-c(6,11,14,20,25,31)+58
EnjoymentKolommen2 <- c(1,5,8,12,22,28,35,39,46,54)+58
HopelessnessKolommen2<-c(40,42,44,49,52,56)+58
PrideKolommen2<-c(18,19,32,36,58,59)+58
ShameKolommen2<-c(10,16,27,33,37,50,57,60)+58

meanAnger= data.frame(NA)
meanBoredom = data.frame(NA)
meanEnjoyment = data.frame(NA)
meanHopelessness = data.frame(NA)
meanPride = data.frame(NA)
meanShame = data.frame(NA)


for (i in 1:207){
    meanAnger[i] = mean(as.numeric(vragenlijst1[i, AngerKolommen]), na.rm = TRUE)-1
    meanBoredom[i] = mean(as.numeric(vragenlijst1[i, BoredomKolommen]), na.rm = TRUE)-1
    meanEnjoyment[i] = mean(as.numeric(vragenlijst1[i, EnjoymentKolommen]), na.rm = TRUE)-
    meanHopelessness[i] = mean(as.numeric(vragenlijst1[i, HopelessnessKolommen]), na.rm =
    meanPride[i] =mean(as.numeric(vragenlijst1[i, PrideKolommen]), na.rm = TRUE)-1
    meanShame[i] = mean(as.numeric(vragenlijst1[i, ShameKolommen]), na.rm = TRUE)-1
```

```
  }

nmmr=0
MatrixEmotionsGrades = matrix(NA, nrow = 80, ncol = 9)
for (k in 1:114){
  for (l in 1:207){
    #als de studentennummers overeenkomen
    if (DataCijfers[k,2] == vragenlijst1[l,2]){
      #voegen we ze in een nieuwe dataset
      nmmr <- nmmr +1
      MatrixEmotionsGrades[nmmr, 1] = DataCijfers[k,2] #stopt leerlingnummer in kolom 1
      MatrixEmotionsGrades[nmmr, 2] = DataCijfers[k,9] #stop alle cijfers van TT1 in kolom 2
      MatrixEmotionsGrades[nmmr,3] = as.character(meananxiety[l]) #gemiddelde anxiety op kolom
      MatrixEmotionsGrades[nmmr,4] = as.character(meanAnger[l])
      MatrixEmotionsGrades[nmmr,5] = as.character(meanBoredom[l])
      MatrixEmotionsGrades[nmmr,6] = as.character(meanEnjoyment[l])
      MatrixEmotionsGrades[nmmr,7] = as.character(meanHopelessness[l])
      MatrixEmotionsGrades[nmmr,8] = as.character(meanPride[l])
      MatrixEmotionsGrades[nmmr,9] = as.character(meanShame[l])
        }
      }
    }


Anxiety <- MatrixEmotionsGrades[,3]
Anger <- MatrixEmotionsGrades[,4]
Boredom <- MatrixEmotionsGrades[,5]
Enjoyment <- MatrixEmotionsGrades[,6]
Hopelessness <- MatrixEmotionsGrades[,7]
Pride <- MatrixEmotionsGrades[,8]
Shame <- MatrixEmotionsGrades[,9]
GradeEmotions <- MatrixEmotionsGrades[,2]

Anxiety <- as.numeric(Anxiety[!is.na(Anxiety)])
Anger <- as.numeric(Anger[!is.na(Anger)])
Boredom <- as.numeric(Boredom[!is.na(Boredom)])
Enjoyment <- as.numeric(Enjoyment[!is.na(Enjoyment)])
Hopelessness <- as.numeric(Hopelessness[!is.na(Hopelessness)])
Pride <- as.numeric(Pride[!is.na(Pride)])
Shame<- as.numeric(Shame[!is.na(Shame)])
GradeEmotions <- as.numeric(GradeEmotions[!is.na(GradeEmotions)])

QR_ALL <- rq(GradeEmotions ~ Anxiety + Anger + Boredom+Enjoyment+Hopelessness+Pride+Shame , dat
QR.ALL <- summary(QR_ALL, se = "boot", bsmethod = "wild")
plot(QR.ALL)

MatCorrelation <- data.frame(NA)
MatCorrelation[1:80,1] <- as.numeric(MatrixEmotionsGrades[,3])
MatCorrelation[1:80,2] <-as.numeric(MatrixEmotionsGrades[,4])
MatCorrelation[1:80,3] <- as.numeric(MatrixEmotionsGrades[,5])
MatCorrelation[1:80,4] <-as.numeric(MatrixEmotionsGrades[,6])
MatCorrelation[1:80,5] <- as.numeric(MatrixEmotionsGrades[,7])
MatCorrelation[1:80,6] <-as.numeric(MatrixEmotionsGrades[,8])
MatCorrelation[1:80,7] <-as.numeric(MatrixEmotionsGrades[,9])
colnames(MatCorrelation) <-c("Anxiety","Anger","Boredom","Enjoyment","Hopelessness","Pride", "S

corrplot(cor(as.matrix(MatCorrelation),use="pairwise.complete.obs"), method = "number")

QR_eliminate <- rq(GradeEmotions ~ Anxiety +Boredom+Enjoyment+Pride, data = as.data.frame(Matri
QR.el <- summary(QR_eliminate, se = "boot")
plot(QR.el)
```

```
#####################################################33333
#SEX AND ALL EMOTIONS
########################################################

meanAnger2= data.frame(NA)
meanBoredom2 = data.frame(NA)
meanEnjoyment2= data.frame(NA)
meanHopelessness2 = data.frame(NA)
meanPride2 = data.frame(NA)
meanShame2 = data.frame(NA)


for (i in 1:121){
  meanAnger2[i] = mean(as.numeric(vragenlijst2[i, AngerKolommen2]), na.rm = TRUE)-1
  meanBoredom2[i] = mean(as.numeric(vragenlijst2[i, BoredomKolommen2]), na.rm = TRUE)-1
  meanEnjoyment2[i] = mean(as.numeric(vragenlijst2[i, EnjoymentKolommen2]), na.rm = TRUE)-
  meanHopelessness2[i] = mean(as.numeric(vragenlijst2[i, HopelessnessKolommen2]), na.rm =
  meanPride2[i] =mean(as.numeric(vragenlijst2[i, PrideKolommen2]), na.rm = TRUE)-1
  meanShame2[i] = mean(as.numeric(vragenlijst2[i, ShameKolommen2]), na.rm = TRUE)-1
}

nmmr=0
MatrixEmotionsGrades2 = matrix(NA, nrow = 80, ncol = 10)
for (k in 1:114){
  for (l in 1:121){
    #als de studentennummers overeenkomen
    if (DataCijfers[k,2] == vragenlijst2[l,2]){
      #voegen we ze in een nieuwe dataset
      nmmr <- nmmr +1
      if (is.na(as.numeric(meananxiety2[l])) == FALSE){
      MatrixEmotionsGrades2[nmmr, 1] = DataCijfers[k,2] #stopt leerlingnummer in kolom 1
      MatrixEmotionsGrades2[nmmr, 2] = DataCijfers[k,9] #stop alle cijfers van TT1 in kolo
      MatrixEmotionsGrades2[nmmr,3] = as.character(meananxiety2[l]) #gemiddelde anxiety op
      MatrixEmotionsGrades2[nmmr,4] = as.character(meanAnger2[l])
      MatrixEmotionsGrades2[nmmr,5] = as.character(meanBoredom2[l])
      MatrixEmotionsGrades2[nmmr,6] = as.character(meanEnjoyment2[l])
      MatrixEmotionsGrades2[nmmr,7] = as.character(meanHopelessness2[l])
      MatrixEmotionsGrades2[nmmr,8] = as.character(meanPride2[l])
      MatrixEmotionsGrades2[nmmr,9] = as.character(meanShame2[l])
      MatrixEmotionsGrades2[nmmr,10] = as.numeric(vragenlijst2[l,6]) #sex op kolom 3
    }
  }
}
}

Anxiety22 <- MatrixEmotionsGrades2[,3]
Anger2 <- MatrixEmotionsGrades2[,4]
Boredom2 <- MatrixEmotionsGrades2[,5]
Enjoyment2 <- MatrixEmotionsGrades2[,6]
Hopelessness2 <- MatrixEmotionsGrades2[,7]
Pride2 <- MatrixEmotionsGrades2[,8]
Shame2 <- MatrixEmotionsGrades2[,9]
Gender <- MatrixEmotionsGrades2[,10]
GradeEmotions2 <- MatrixEmotionsGrades2[,2]

Anxiety22 <- as.numeric(Anxiety22[!is.na(Anxiety22)])
Anger2 <- as.numeric(Anger2[!is.na(Anger2)])
Boredom2 <- as.numeric(Boredom2[!is.na(Boredom2)])
Enjoyment2 <- as.numeric(Enjoyment2[!is.na(Enjoyment2)])
Hopelessness2 <- as.numeric(Hopelessness2[!is.na(Hopelessness2)])
```

```
Pride2 <- as.numeric(Pride2[!is.na(Pride2)])
Shame2<- as.numeric(Shame2[!is.na(Shame2)])
Gender<- as.factor(Gender[!is.na(Gender)])
GradeEmotions2 <- as.numeric(GradeEmotions2[!is.na(GradeEmotions2)])


QR_ALL2 <- rq(GradeEmotions2 ~ Anxiety22+Anger2+Boredom2+Hopelessness2+Pride2+Shame2+Gender, da
QR.ALL2 <- summary(QR_ALL2, se = "boot", bsmethod = "wild")
plot(QR.ALL2)

QR_GENDER <- rq(Anxiety22 ~ Gender, data = as.data.frame(MatrixEmotionsGrades2), tau = seq(0.05
QR.GEN <- summary(QR_GENDER, se = "boot")
plot(QR.GEN, main = c("Intercept", "Influence␣of␣Gender␣on␣different␣quantiles␣of␣Shame" ))
```

B
Questionnaire

Questionnaire_week38_2022

Start of Block: Opening

Intro Beste student,

Heel erg bedankt voor het meedoen aan dit onderzoek. In dit onderzoek kijken we naar zelf gereguleerd leren en emoties die bij leren en presteren komen kijken. Hiermee kunnen we begrijpen hoe studenten leren, en hoe dit leren ondersteund kan worden.

De vragenlijst bestaat uit drie delen waarmee naar verschillende onderdelen van het leren wordt gekeken. Het eerste deel gaat voornamelijk over wat voor leer-strategieën je gebruikt bij dit vak en welke studievaardigheden jij denkt nodig te hebben. Het tweede en derde deel gaan over verschillende emoties die bij je opkomen voor, tijdens en na het college, het leren voor het tentamen, en het tentamen zelf.

In totaal duurt de vragenlijst ongeveer 15 minuten. Er zijn geen goede of foute antwoorden, het gaat om jouw ervaring. Vragen of opmerkingen zijn altijd welkom. **Mail naar Andreas de Hartog** (A.J.deHartog@student.tudelft.nl)

Onder de deelnemers worden 10 cadeaubonnen van 10 euro verloot. Indien je in de prijzen valt krijg je daarvan bericht!

Page Break

Q37 Ik verklaar hierbij dat ik 17 jaar of ouder ben en toestemming geef om deel te nemen aan het onderzoek genaamd **Zelf-gereguleerd leren en academische emoties in wiskunde service onderwijs**.

Toesteming
Ik heb de volgende voorwaarden gelezen, begrepen en geaccepteerd: Het doel van dit onderzoek is om te bepalen hoe studenten zelf-gereguleerd leren en academische emoties ervaren in wiskunde service onderwijs. Deelname aan dit onderzoek helpt met het beter begrijpen van hoe studenten studeren en hoe studeren kan worden ondersteund. De vragenlijst duurt ongeveer 15 minuten. Mij wordt een aantal vragen gesteld. Alle vragen zijn multiple-choice en er zijn geen goede of foute antwoorden. Het gaat om het antwoord dat het meest van toepassing is. De onderzoekers zullen alle vragen beantwoorden die ik nu of gedurende het onderzoek heb.

Ik heb de volgende data voorwaarden gelezen, begrepen en geaccepteerd: Mijn deelname is vrijwillig. Ik heb de optie om op ieder moment met het onderzoek te stoppen zonder opgaaf van redenen. De data zal vertrouwelijk en anoniem behandeld worden. Het resultaat van het onderzoek kan niet naar mij als individu worden gelinkt. De verzamelde data zullen alleen voor onderzoeksdoeleinden worden gebruikt en alleen het onderzoeksteam heeft toegang tot de data. De resultaten mogen, in verband gebracht worden met studiegegevens zoals cijfers en studievoortgang. De data wordt voor 10 jaar lang bewaard, in overeenkomst met de VSNU gedragscode.

Handtekening Zet hier je handtekening als je het met de vorige voorwaarden eens bent.

Page Break

Gegevens Voordat we beginnen willen we graag wat gegevens van je vragen.

- ◯ Ik (naam) (1) ____
- ◯ Studentnummer (2) ____
- ◯ Studie (welke studie volg je?) (8) ____
- ◯ Leeftijd (5) ____
- ◯ Gemiddelde eindcijfer voor wiskunde bij mijn vooropleiding (middelbare school of HBO) (3) ____

Q43 Geslacht

- ◯ Man (1)
- ◯ Vrouw (2)
- ◯ Non-binair (3)
- ◯ Wil ik liever niet zeggen (4)
- ◯ Anders, namelijk: (5) ____

Q42 Hoeveel uur besteed je gemiddeld per week aan de wiskundecursus?

- ◯ 10 uur (1)
- ◯ 10-20 uur (2)
- ◯ 20-30 uur (3)
- ◯ 30-40 uur (4)
- ◯ 40-50 uur (5)
- ◯ meer dan 50 uur (6)

Q35 Als er nog vragen zijn in verband met dit onderzoek, zouden wij het heel erg waarderen als je contact met ons opneemt. Stuur in dat geval een email naar Andreas de Hartog (A.J.deHartog@student.tudelft.nl).

End of Block: Opening

Start of Block: MSLQ

MSLQ **Leerstrategieën (Deel 1)**

De volgende vragen gaan over jouw leerstrategieën en studievaardigheden voor dit vak. Er zijn geen goede of foute antwoorden. Beantwoord de vraag over hoe je in de klas studeert zo

precies mogelijk. Als je denkt dat de stelling heel erg op jou van toepassing is, kies 7. Als de stelling helemaal niet op jou van toepassing is, kies 1.

	Helemaal <u>niet</u> op mij van toepassing 1. (1)	2. (2)	3. (3)	4. (4)	5. (5)	6. (6)	**Helemaal op mij van toepassing** 7. (7)
1. Wanneer ik de hoofdstukken bestudeer, probeer ik eerst de hoofdzaken te bepalen om zo overzicht te krijgen. (1)	○	○	○	○	○	○	○
2. Tijdens het college mis ik vaak belangrijke punten omdat ik met mijn gedachten ergens anders ben. (2)	○	○	○	○	○	○	○
3. Als ik leer voor een toets, leg ik wat ik leer vaak uit aan vrienden en/of klasgenoten. (3)	○	○	○	○	○	○	○
4. Meestal studeer ik op een plaats waar ik mij kan concentreren. (4)	○	○	○	○	○	○	○

5. Wanneer ik voor dit vak aan het studeren ben, bedenk ik vragen om mijn aandacht erbij te houden. (5)	○	○	○	○	○	○	○
6. Tijdens het leren voor een toets ben ik vaak zo verveeld, dat ik eerder stop met leren dan dat ik eigenlijk gepland had. (6)	○	○	○	○	○	○	○
7. Ik stel mezelf vragen over dat wat ik bestudeer of hoor tijdens de colleges, om te bepalen of ik geloof wat ik bestudeerd heb of wat er gezegd wordt. (7)	○	○	○	○	○	○	○
8. Wanneer ik studeer voor een toets herhaal ik steeds hardop alle feiten. (8)	○	○	○	○	○	○	○

9. Ook al heb ik moeite met dit vak, ik blijf het alleen proberen, zonder de hulp van anderen. (9)	○	○	○	○	○	○	○
10. Wanneer ik tijdens het studeren voor dit vak iets niet begrijp, ga ik het nog een keer bestuderen en probeer ik eruit te komen. (10)	○	○	○	○	○	○	○

Page Break

Q44 **Leerstrategieën (Deel 1)**

De volgende vragen gaan over jouw leerstrategieën en studievaardigheden voor dit vak. Er zijn geen goede of foute antwoorden. Beantwoord de vraag over hoe je in de klas studeert zo

precies mogelijk. Als je denkt dat de stelling heel erg op jou van toepassing is, kies 7. Als de stelling helemaal niet op jou van toepassing is, kies 1.

	Helemaal niet op mij van toepassing 1. (1)	2. (2)	3. (3)	4. (4)	5. (5)	6. (6)	**Helemaal op mij van toepassing** 7. (7)
11. Wanneer ik leer voor een toets bekijk ik de boeken en mijn aantekeningen en ga ik op zoek naar de belangrijkste onderdelen. (11)	○	○	○	○	○	○	○
12. Ik maak efficiënt gebruik van mijn tijd die ik heb om te studeren. (12)	○	○	○	○	○	○	○
13. Wanneer ik een hoofdstuk bestudeer en dat hoofdstuk niet begrijp, verander ik mijn manier van studeren. (13)	○	○	○	○	○	○	○
14. Ik werk samen met studiegenoten om opdrachten voor dit vak af te maken. (14)	○	○	○	○	○	○	○

15. Wanneer ik leer voor een toets bestudeer ik mijn aantekeningen en de boeken keer-op-keer. (15)	○	○	○	○	○	○	○
16. Wanneer er bij een bijeenkomst gesproken wordt over een onderwerp, probeer ik voor mezelf helder te krijgen of er goed ondersteunend bewijs voor is. (16)	○	○	○	○	○	○	○
17. Ik werk hard voor dit vak ook al vind ik het vak helemaal niet leuk. (17)	○	○	○	○	○	○	○
18. Ik maak lijstjes, schema's etc. om zo overzicht te krijgen over datgene wat ik moet maken en doen voor het vak. (18)	○	○	○	○	○	○	○
19. Wanneer ik leer voor de toets maak ik tijd vrij om met studiegenoten het lesmateriaal te bespreken. (19)	○	○	○	○	○	○	○

20. Ik probeer mijn eigen ideeën te vormen over datgene wat ik leer tijdens het college. (20)	○	○	○	○	○	○	○

Page Break

Q45 **Leerstrategieën (Deel 1)**

De volgende vragen gaan over jouw leerstrategieën en studievaardigheden voor dit vak. Er zijn geen goede of foute antwoorden. Beantwoord de vraag over hoe je in de klas studeert zo

precies mogelijk. Als je denkt dat de stelling heel erg op jou van toepassing is, kies 7. Als de stelling helemaal niet op jou van toepassing is, kies 1.

	Helemaal <u>niet</u> op mij van toepassing 1. (1)	2. (2)	3. (3)	4. (4)	5. (5)	6. (6)	**Helemaal op mij van toepassing** 7. (7)
21. Ik vind het moeilijk om me aan een planning te houden. (21)	○	○	○	○	○	○	○
22. Wanneer ik leer voor de toets maak ik gebruik van meerdere bronnen, zoals de colleges, boeken/readers en eigen aantekeningen. (22)	○	○	○	○	○	○	○
23. Voordat ik een nieuw hoofdstuk ga bestuderen, bekijk ik eerst hoe het hoofdstuk is opgebouwd. (23)	○	○	○	○	○	○	○
24. Ik stel mezelf vragen om er zeker van te zijn dat ik het collegestof begrijp. (24)	○	○	○	○	○	○	○
25. Ik pas mijn leerstijl aan de eisen van het college en de manier van lesgeven van de docent aan. (25)	○	○	○	○	○	○	○

26. Het komt geregeld voor dat ik iets heb bestudeerd voor dit vak en dat ik niet echt begrijp waar het over gaat. (52)	○	○	○	○	○	○	○
27. Wanneer ik onderdelen niet begrijp, vraag ik de docent om uitleg. (53)	○	○	○	○	○	○	○
28. Kern begrippen helpen mij om belangrijke onderdelen te onthouden. (54)	○	○	○	○	○	○	○
29. Wanneer een opdracht moeilijk is, stop ik ermee of ik doe alleen de makkelijke delen. (55)	○	○	○	○	○	○	○
30. Wanner ik leer voor een toets denk ik eerst na over het onderwerp en bepaal ik wat ik moet leren, in plaats van gewoon het hoofdstuk te bestuderen. (56)	○	○	○	○	○	○	○

Page Break

Q41 **Leerstrategieën (Deel 2)**

De volgende vragen gaan over jouw leerstrategieën en studievaardigheden voor dit vak. Er zijn geen goede of foute antwoorden. Beantwoord de vraag over hoe je in de klas studeert zo

precies mogelijk. Als je denkt dat de stelling heel erg op jou van toepassing is, kies 7. Als de stelling helemaal niet op jou van toepassing is, kies 1.

	Helemaal niet op mij van toepassing 1. (1)	2. (2)	3. (3)	4. (4)	5. (5)	6. (6)	**Helemaal op mij van toepassing** 7. (7)
31. Ik probeer de onderwerpen van dit vak te verbinden met andere vakken. (31)	○	○	○	○	○	○	○
32. Wanneer ik studeer voor een toets maak ik vanuit mijn aantekeningen een lijst met de belangrijkste onderdelen. (32)	○	○	○	○	○	○	○
33. Wanneer ik studeer voor dit vak probeer ik te bedenken wat ik eigenlijk al van het onderwerp weet. (33)	○	○	○	○	○	○	○
34. Ik maak mijn huiswerk vaak op dezelfde plek. (34)	○	○	○	○	○	○	○
35. Ik ben graag bezig met mijn eigen ideeën over datgene wat ik leer in het college. (35)	○	○	○	○	○	○	○

36. Wanneer ik leer voor een toets maak ik een korte samenvatting van mijn aantekeningen en de hoofdstukken uit de boeken. (36)	○	○	○	○	○	○	○
37. Wanneer ik iets niet begrijp vraag ik een studiegenoot om uitleg. (37)	○	○	○	○	○	○	○
38. Ik probeer onderwerpen uit meerdere colleges aan elkaar te verbinden. (38)	○	○	○	○	○	○	○
39. Ik zorg dat ik bij blijf met de colleges van dit vak. (39)	○	○	○	○	○	○	○
40. Bij een probleem denk ik altijd na over meerdere oplossingen. (40)	○	○	○	○	○	○	○

Page Break

Q46 **Leerstrategieën (Deel 2)**

De volgende vragen gaan over jouw leerstrategieën en studievaardigheden voor dit vak. Er zijn geen goede of foute antwoorden. Beantwoord de vraag over hoe je in de klas studeert zo

precies mogelijk. Als je denkt dat de stelling heel erg op jou van toepassing is, kies 7. Als de stelling helemaal niet op jou van toepassing is, kies 1.

	Helemaal <u>niet</u> op mij van toepassing 1. (1)	2. (2)	3. (3)	4. (4)	5. (5)	6. (6)	Helemaal op mij van toepassing 7. (7)
41. Ik maak lijstjes van de belangrijkste punten en deze leer ik uit mijn hoofd. (41)	o	o	o	o	o	o	o
42. Ik ben regelmatig aanwezig bij de colleges. (42)	o	o	o	o	o	o	o
43. Ook al is de lesstof saai, het lukt me toch om de opdrachten af te maken. (43)	o	o	o	o	o	o	o
44. Ik ga op zoek naar studiegenoten die ik om hulp kan vragen wanneer dit nodig mocht zijn. (44)	o	o	o	o	o	o	o
45. Als ik leer voor een toets zoek ik uit welke begrippen ik nog niet goed begrepen heb. (45)	o	o	o	o	o	o	o
46. Ik merk dat ik niet veel tijd besteed aan dit vak, omdat ik bezig ben met andere zaken. (46)	o	o	o	o	o	o	o
47. Ik maak regelmatig einddoelen voor mezelf om het maken van mijn huiswerk een richting te geven. (47)	o	o	o	o	o	o	o
48. Als ik tijdens het college onduidelijke aantekeningen maak, zoek ik het na het college direct uit. (48)	o	o	o	o	o	o	o
49. Ik heb bijna nooit tijd om mijn aantekeningen en boeken door te nemen voor een toets. (49)	o	o	o	o	o	o	o

50. Wat ik heb bestudeerd pas ik toe bij samenwerkingsopdrachten. (50)	○	○	○	○	○	○	○

End of Block: MSLQ

Start of Block: AEQ-M

Intro
Emoties bij wiskunde
De komende vragenlijst gaat over jouw gevoelens en gedachten over het vak wiskunde. Er zijn geen goede of foute antwoorden – we willen weten hoe jij je voelt en hoe jij denkt over je ervaringen bij wiskunde. We zijn geïnteresseerd in jouw persoonlijke mening, dus wees eerlijk en denk niet te lang na bij het beantwoorden van de vragen.

De vragenlijst bestaat uit 3 korte delen.

Page Break

Emoties bij de les **Deel I - Emoties bij het college**

Het volgen van wiskundecolleges kan verschillende gevoelens oproepen. Dit deel van de vragenlijst gaat over emoties die jij kan ervaren gedurende het wiskundecollege. Denk voor het beantwoorden van de volgende vragen aan typische situaties die jij ervaren hebt in jouw wiskundecolleges.

Vóór de les **Vóór het college**

De volgende vragen gaan over gevoelens die jij kan ervaren vóór het wiskundecollege. Geef alsjeblieft aan hoe jij je meestal voelt voordat jij naar het wiskundecollege gaat.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
1. Ik kijk uit naar de wiskundecolleges. (1)	○	○	○	○	○
2. Ik word nerveus wanneer ik aan de wiskundecolleges denk. (2)	○	○	○	○	○
3. Ik word misselijk wanneer ik aan mijn wiskundecolleges denk. (3)	○	○	○	○	○
4. Wiskunde maakt me zo bang dat ik liever niet naar de colleges zou gaan. (4)	○	○	○	○	○

Tijdens de les **Tijdens het college**

De volgende vragen gaan over gevoelens die jij kan ervaren tijdens het wiskundecollege. Geef alsjeblieft aan hoe jij je meestal voelt tijdens het wiskundecollege.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
5. Ik heb plezier in de colleges. (1)	◯	◯	◯	◯	◯
6. Ik kan me niet concentreren omdat ik me zo verveel. (11)	◯	◯	◯	◯	◯
7. Ik maak me zorgen dat de stof te moeilijk voor mij is. (2)	◯	◯	◯	◯	◯
8. Ik vind de stof die we in de colleges behandelen zo fascinerend, dat ik veel plezier heb tijdens colleges. (3)	◯	◯	◯	◯	◯
9. Ik voel me geërgerd tijdens de colleges. (7)	◯	◯	◯	◯	◯
10. Wanneer ik iets zeg tijdens de colleges, merk ik dat mijn gezicht rood wordt. (4)	◯	◯	◯	◯	◯
11. Ik vind de colleges saai (12)	◯	◯	◯	◯	◯

12. Ik vind de colleges zo leuk dat ik erg gemotiveerd ben om mee te doen. (5)	○	○	○	○	○
13. Ik ben zo boos tijdens de colleges dat ik graag weg zou lopen. (8)	○	○	○	○	○
14. Ik ben zo verveeld dat ik niet wakker kan blijven. (13)	○	○	○	○	○
15. Ik word boos omdat de stof in het college zo moeilijk is. (9)	○	○	○	○	○
16. Wanneer ik iets zeg tijdens het college heb ik het gevoel dat ik mezelf voor schut zet. (6)	○	○	○	○	○
17. Ik word geïrriteerd door de colleges. (10)	○	○	○	○	○

Q38 **Na het college**
De volgende vragen gaan over gevoelens die jij kan ervaren na het wiskundecollege. Geef alsjeblieft aan hoe jij je meestal voelt na het wiskundecollege.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
18. Ik kan denk ik trots zijn op mijn wiskundekennis. (1)	○	○	○	○	○
19. Ik ben trots op mijn bijdragen tijdens de colleges. (2)	○	○	○	○	○

Page Break

Emoties bij het lere **Deel II - Emoties bij het leren**
Leren en huiswerk maken voor wiskunde kan verschillende gevoelens oproepen. Dit deel van de vragenlijst gaat over emoties die jij kan ervaren wanneer je leert en huiswerk maakt voor wiskunde. Denk voor het beantwoorden van de volgende vragen, aan typische situaties die jij ervaren hebt tijdens het leren of maken van huiswerk voor wiskunde.

Vóór het leren **Vóór het leren**
De volgende vragen hebben betrekking op gevoelens die jij kan ervaren vóór het leren en huiswerk maken voor wiskunde. Geef alsjeblieft aan hoe jij je meestal voelt voordat je begint met leren en het maken van huiswerk voor wiskunde.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
20. Als ik denk aan mijn wiskundehuiswerk, begin ik me al te vervelen. (2)	○	○	○	○	○
21. Ik ben zo bang voor mijn wiskundehuiswerk dat ik er liever niet aan zou beginnen. (1)	○	○	○	○	○

Tijdens het leren **Tijdens het leren**
De volgende vragen hebben betrekking op gevoelens die jij kan ervaren tijdens het leren en

huiswerk maken voor wiskunde. Geef alsjeblieft aan hoe jij je meestal voelt tijdens het leren en maken van opdrachten voor wiskunde.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
22. Wanneer ik mijn wiskundehuiswerk maak, ben ik in een goede bui. (1)	○	○	○	○	○
23. Ik begin te zweten, omdat ik me zorgen maak dat ik mijn huiswerk niet op tijd af krijg. (2)	○	○	○	○	○
24. Het maken van wiskundehuiswerk maakt me boos. (10)	○	○	○	○	○
25. Ik verveel me zo erg dat ik geen zin meer heb om te studeren. (13)	○	○	○	○	○
26. Ik ben gespannen en zenuwachtig. (3)	○	○	○	○	○
27. Wanneer ik opdrachten bespreek met mijn studiegenoten vermijd ik oogcontact. (4)	○	○	○	○	○
28. Ik ben blij dat ik de stof begrijp. (5)	○	○	○	○	○
29. Ik word boos omdat mijn wiskundehuiswerk zo veel tijd in beslag neemt. (11)	○	○	○	○	○

30. Ik maak me zorgen over of ik ooit in staat zal zijn om de stof helemaal te begrijpen. (6)	○	○	○	○	○
31. Ik verveel me dood bij het maken van wiskundehuiswerk. (14)	○	○	○	○	○
32. Ik ben erg gemotiveerd, want ik wil trots zijn op mijn resultaten in wiskunde. (9)	○	○	○	○	○
33. Wanneer ik iets niet begrijp bij het huiswerk, wil ik dit aan niemand vertellen. (7)	○	○	○	○	○
34. Ik ben zo boos dat ik het huiswerk wel in de prullenbak zou willen gooien. (12)	○	○	○	○	○
35. Ik vind het zo leuk om mijn wiskundehuiswerk te doen dat ik gemotiveerd ben om extra opdrachten te maken. (8)	○	○	○	○	○

Na het leren **Na het leren**

De volgende vragen hebben betrekking op gevoelens die jij kan ervaren na het leren en maken van huiswerk voor wiskunde. Geef alsjeblieft aan hoe jij je meestal voelt nadat je hebt geleerd of opgaven hebt gemaakt voor wiskunde.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
36. Nadat ik mijn wiskundehuiswerk heb gemaakt ben ik trots op mijzelf. (2)	o	o	o	o	o
37. Ik schaam me voor mijn gebrek aan kennis van de wiskunde. (1)	o	o	o	o	o

Page Break

Emoties bij toetsen **Deel III - Emoties bij toetsen/ tentamens**
Toetsen en tentamens voor wiskunde kunnen verschillende gevoelens oproepen. Dit deel van de vragenlijst gaat over emoties die jij kan ervaren wanneer je wiskundetoetsen of -tentamens maakt. Denk voor het beantwoorden van de volgende vragen aan typische situaties die jij tijdens het maken van wiskundetoetsen of -tentamens ervaren hebt.

Vóór de toets/ten **Vóór de toets/ tentamen**
De volgende vragen hebben betrekking op gevoelens die jij kan ervaren vóór het maken van

een wiskundetoets of -tentamen. Geef alsjeblieft aan hoe jij je meestal voelt voor het maken van een wiskundetoets of -tentamen.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
38. Ik ben erg nerveus. (1)	○	○	○	○	○
39. Ik kijk uit naar een hoog cijfer, dus ik studeer hard voor de toets. (2)	○	○	○	○	○
40. Ik voel me verdrietig. (6)	○	○	○	○	○
41. Zelfs voordat de toets begint, maak ik me zorgen dat ik het niet zal halen. (3)	○	○	○	○	○
42. Ik denk alsmaar dat ik de stof niet begrijp. (7)	○	○	○	○	○
43. Wanneer er een toets komt, voel ik me ziek worden. (4)	○	○	○	○	○
44. Ik denk alsmaar dat ik nooit goede cijfers zal halen bij wiskunde. (8)	○	○	○	○	○
45. Ik ben zo gespannen dat ik de toets liever niet maak. (5)	○	○	○	○	○

Tijdens de toets/ten **Tijdens de toets/ tentamen**
De volgende vragen hebben betrekking op gevoelens die jij kan ervaren tijdens het maken van

een toets of tentamen voor wiskunde. Geef alsjeblieft aan hoe jij je meestal voelt tijdens een wiskundetoets of tentamen.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
46. Ik vind het leuk om wiskundetoetsen te maken. (1)	○	○	○	○	○
47. Wanneer ik een wiskundetoets maak, voel ik me gespannen en nerveus. (2)	○	○	○	○	○
48. Ik erger me er aan dat de docent zulke moeilijke vragen stelt. (8)	○	○	○	○	○
49. Ik voel me hopeloos tijdens de wiskunde toets. (10)	○	○	○	○	○
50. Ik schaam me dat ik de vragen van de docent niet goed kan beantwoorden. (3)	○	○	○	○	○
51. Ik ben zo gespannen dat ik mij niet volledig kan concentreren. (4)	○	○	○	○	○
52. Ik zou het liever opgeven. (11)	○	○	○	○	○
53. Ik ben zo boos dat ik de toets in stukjes zou willen scheuren. (9)	○	○	○	○	○

54. Ik denk dat het geweldig gaat. (5)	○	○	○	○	○
55. Tijdens het maken van de toets maak ik me zorgen dat ik een slecht cijfer zal halen. (6)	○	○	○	○	○
56. Ik heb geen energie. (12)	○	○	○	○	○
57. Ik schaam me zo voor mijn prestaties op de wiskundetoets dat ik begin te zweten. (7)	○	○	○	○	○

Na de toets/ten **Na de toets/ tentamen**

De volgende vragen hebben betrekking op gevoelens die jij kan ervaren na het maken van een toets of tentamen voor wiskunde. Geef alsjeblieft aan hoe jij je meestal voelt na het maken van een wiskundetoets of tentamen.

	Helemaal mee oneens 1. (1)	2. (2)	3. (3)	4. (4)	**Helemaal mee eens** 5. (5)
58. Na een wiskundetoets ben ik trots op mijzelf. (2)	○	○	○	○	○
59. Ik ben trots op hoe goed ik de wiskundetoets heb gemaakt. (3)	○	○	○	○	○
60. Na het maken van een wiskundetoets voel ik me beschaamd. (4)	○	○	○	○	○

End of Block: AEQ-M

Start of Block: EES

Q29

Aanvullende emoties bij wiskunde

Tot slot zijn we geïnteresseerd in enkele aanvullende emoties/ gevoelens die je ervaart tijdens het studeren van wiskunde. Geef voor elke emotie aan **hoe sterk** je deze ervaart, door het passende nummer in te vullen.

	Helemaal niet (1)	Een beetje (2)	Gemiddeld (3)	Sterk (4)	Heel sterk (5)
1. Nieuwsgierig (1)	○	○	○	○	○
2. Verveeld (2)	○	○	○	○	○
3. Vertwijfeld (3)	○	○	○	○	○
4. Verrast (4)	○	○	○	○	○
5. Geïnteresseerd (5)	○	○	○	○	○
6. Ongerust (6)	○	○	○	○	○
7. Gefrustreerd (7)	○	○	○	○	○
8. Onderzoekend (8)	○	○	○	○	○
9. Saai (9)	○	○	○	○	○
10. Gefascineerd (10)	○	○	○	○	○
11. Bezorgd (11)	○	○	○	○	○
12. Blij (12)	○	○	○	○	○
13. Verward (13)	○	○	○	○	○
14. Geïrriteerd (14)	○	○	○	○	○
15. Eentonig (15)	○	○	○	○	○

16. Enthousiast (16)	○	○	○	○	○
17. Verbaasd (17)	○	○	○	○	○
18. Ontevreden (18)	○	○	○	○	○
19. Nerveus (19)	○	○	○	○	○
20. Verheugd (20)	○	○	○	○	○
21. Onzeker (21)	○	○	○	○	○
22. Hoopvol (22)	○	○	○	○	○
23. Opgelucht (23)	○	○	○	○	○

End of Block: EES

Bibliography

[1] J.H. Albert and S. Chib. "Bayesian analysis of binary and polychotomous response data". In: (1993) (cit. on p. 14).

[2] D.F. Andrews and C.L. Mallows. "Scale mixtures of normal distributions". In: (1974) (cit. on p. 15).

[3] *Conditional Probability Distribution*. Dec. 2022. URL: `https://en.wikipedia.org/wiki/Conditional_%20probability_%20distribution`.

[4] *Conjugate prior*. Dec. 2022. URL: `https://en.wikipedia.org/wiki/Conjugate_prior` (cit. on p. 11).

[5] Hinkley D.V. Cox D.R. *Theoretical Statistics*. Chapman Hall, 1974, pp. 214–233 (cit. on p. 22).

[6] Russel Davidson and Emmanuel Flachaire. "The wild bootstrap, tamed at last". In: (2008) (cit. on p. 7).

[7] TU Delft. *PRIME*. Dec. 2022. URL: `https://www.tudelft.nl/en/eemcs/the-faculty/departments/applied-mathematics/education/prime/` (cit. on p. 17).

[8] Scott D. Brown Don van Ravenzwaaij Pete Cassey. "A simple introduction to Markov Chain Monte-Carlo sampling". In: (2018) (cit. on p. 12).

[9] Aad van der Vaart Fetsje Bijma Marianne Jonker. *An Introduction to Mathematical Statistics*. Amsterdam University Press, 2016 (cit. on pp. 4, 13).

[10] M. Geraci and M. Bottai. "Quantile regression for longitudinal data using the asymmetric laplace distribution". In: (2007) (cit. on p. 14).

[11] Artur Gramacki. *Nonparametric Kernel Density Estimation and Its Computational Aspect*. Springer, 2018 (cit. on p. 19).

[12] Jorge M UribeMontserrat Guillen. *Quantile regression for cross-sectional and time series data : applications in energy markets using R*. Springer, 2020 (cit. on p. 20).

[13] Yeongjin Gwon Joseph G. Ibrahim Ming-Hui Chen. "The Power Prior: Theory and Applications". In: (2016) (cit. on pp. 11, 12).

[14] Rana A. Moyeed Keming Yu. "Bayesian quantile regression". In: (2001) (cit. on pp. 9–12, 15).

[15] Donald E. Knuth. "Literate Programming". In: *The Computer Journal* 27.2 (1984), pp. 97–111.

[16] Donald E. Knuth. *The TEX Book*. Addison-Wesley Professional, 1986.

[17] R. Koenker and G. Bassett. *Regression Quantiles*. Econometrica, 1978, pp. 33–50 (cit. on pp. 4, 5, 10).

[18] H. Kozumi and G. Kobayashi. "Gibbs sampling methods for Bayesian quantile regression". In: (2011) (cit. on p. 14).

[19] John K. Kruschke. *Doing Bayesian Data Analysis*. Science Direct, 2015 (cit. on p. 9).

[20] Leslie Lamport. *LATEX: a Document Preparation System*. 2nd ed. Massachusetts: Addison Wesley, 1994.

[21] Daniel Sabanés Bové Leonhard Held. *Applied statistical inference: likelihood and Bayes*. Springer, 2014 (cit. on p. 9).

[22] Michael Lesk and Brian Kernighan. "Computer Typesetting of Technical Journals on UNIX". In: *Proceedings of American Federation of Information Processing Societies: 1977 National Computer Conference*. Dallas, Texas, 1977, pp. 879–888.

[23] *Multicollinearity*. Dec. 2022. URL: `https://en.wikipedia.org/wiki/Multicollinearity` (cit. on p. 22).

[24] Yaacov Petscher. "Literate Programming". In: (2016) (cit. on p. 17).

[25] Haithem Taha Mohammed Ali Rahim Alhamzawi. "Bayesian quantile regression for ordinal longitudinal data". In: (2018) (cit. on pp. 13–15).

[26] Keming Yu Rahim Alhamzawi. “Power Prior Elicitation in Bayesian Quantile Regression”. In: (2011) (cit. on pp. 11, 12).

[27] Thomas Goetz Reinhard Pekrun. “Academic Emotions Questionnaire (AEQ)”. In: (2005) (cit. on pp. 1, 17).

[28] Stefania Mignani Silvia Cagnone and Irini Moustaki. “Latent variable models for ordinal data”. In: (2009) (cit. on p. 14).

[29] David Nott Y. Fan Yanan. “Flexible Bayesian Regression Modeling”. In: (2020) (cit. on p. 10).