

# SAA for many N level multilevel models

Welcome to an SAA for fitting many model types developed for Stat-JR v1.0.5

## Input questions

Firstly on this page you will need to specify the dataset required from the list of available datasets.

---

**Which dataset do you wish to use:**

Submit

Next you need to choose many options including the response, estimation method, clustering variables and predictor variables (both continuous and categorical) from the chosen dataset. After choosing these variables the SAA will run and you will see a block of text describing how many observations are to be used at the bottom of this page. The rest of the analysis will appear in pages 2-12.

---

**What estimation method do you want to use:**

IGLS

---

**What is the response variable:**

a\_point

---

**What distribution are you going to assume:**

Normal

---

**Please enter your possible (nested) classifications / levels (lowest first, not including level-1):**

estab,lea

---

**Are there any continuous predictors that need including in all models:**

No

<b>Are there any categorical predictors that need including in all models:</b>	No
<b>Do you want to include any continuous predictors as candidates for inclusion in the models:</b>	Yes
<b>Which continuous predictors do you want to consider:</b>	gcse_tot,gcse_no
<b>Do you want to include any categorical predictors as candidates for inclusion in the models:</b>	Yes
<b>Which categorical predictors do you want to consider:</b>	gender
<b>What selection type do you require:</b>	Forward pass
<b>Do you want to test for random slopes:</b>	Yes
<b>Do you want to test for interactions:</b>	Yes
<b>How do you wish to compare models:</b>	Likelihood Ratio

The Analysis Assistant you are currently using is designed to work on complete datasets only and so as a pre-processing step we have to remove any rows that contain missing data in columns used in the analysis that follows. For now the list of columns to be considered is: a\_point, estab, lea, gcse\_tot, gcse\_no, gender. There are 0 (0.0%) rows that get deleted This results in a dataset of 2166 rows. On the next page we will look at the shape of the response and, in the case of normal responses, decide whether to log transform.

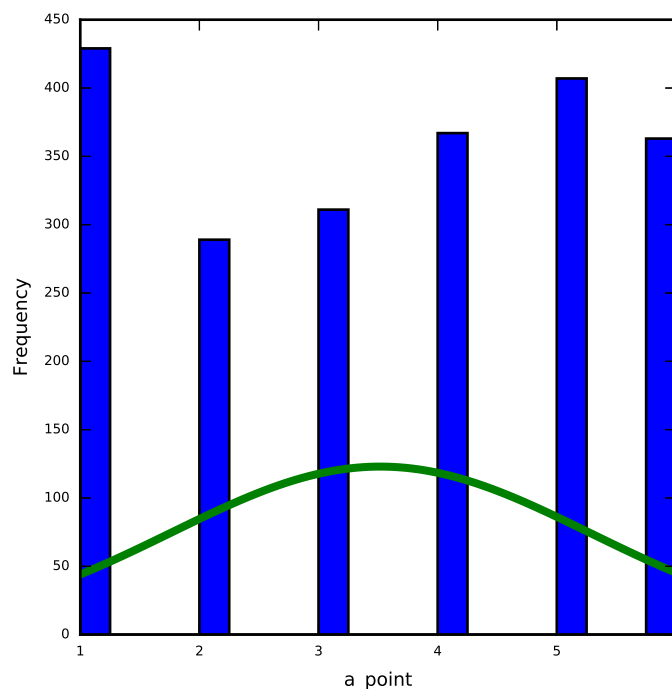
## Exploring the response

We will begin our analysis of the dataset by doing some basic data exploration.

You have chosen `a_point` as your response variable and so a first step is to take a look at this variable and assess its suitability for a normal model. The summary statistics for the variable are in the table below:

	<b>Observations</b>	2166
	<b>Mean</b>	3.518
	<b>Standard Deviation</b>	1.757
	<b>Median</b>	4.0

We also look at a histogram of `a_point` to see if it is approximately normally distributed. Although in modelling the response in terms of a set of predictors it is what is unexplained (the model residuals) that need to be normally distributed, it is still useful to look at the response variable as a very skewed variable will often lead to very skewed residuals.



Here the distribution is reasonably symmetric with skewness value  $-0.085$ .

There are no obvious outliers in `a_point`.

---

**Do you want to log transform the  
response variable?:**

No

## Exploring the predictors individually

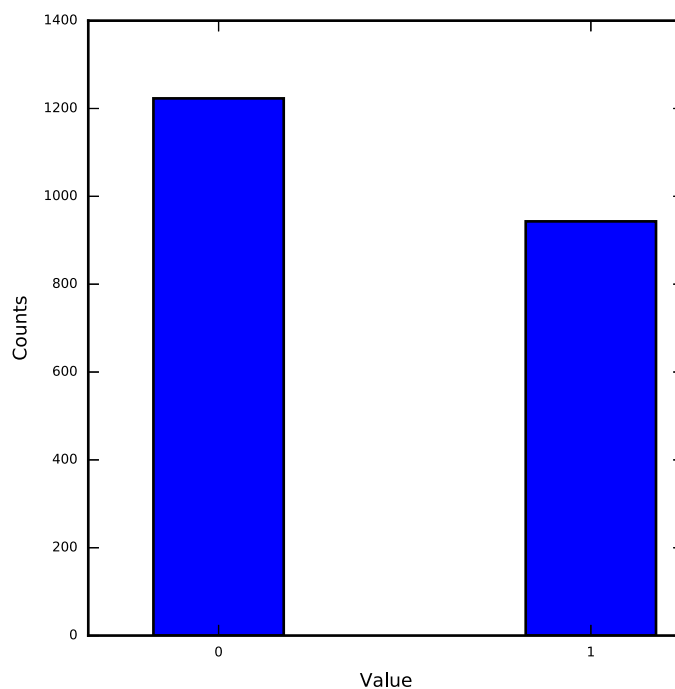
We can also look at each of the predictor variables in turn in isolation.

For categorical predictors we are looking at how common each category is in the dataset. In particular we are checking for rare categories which might cause difficulties in modelling and might therefore be usefully merged with other categories (though this would need to be done outside this SAA).

For predictor gender we see the following:

gender	N	Percentage
0	1223	56.464
1	943	43.536
<b>Total</b>	<b>2166</b>	<b>100</b>

None of the categories of gender have fewer than 5 observations.

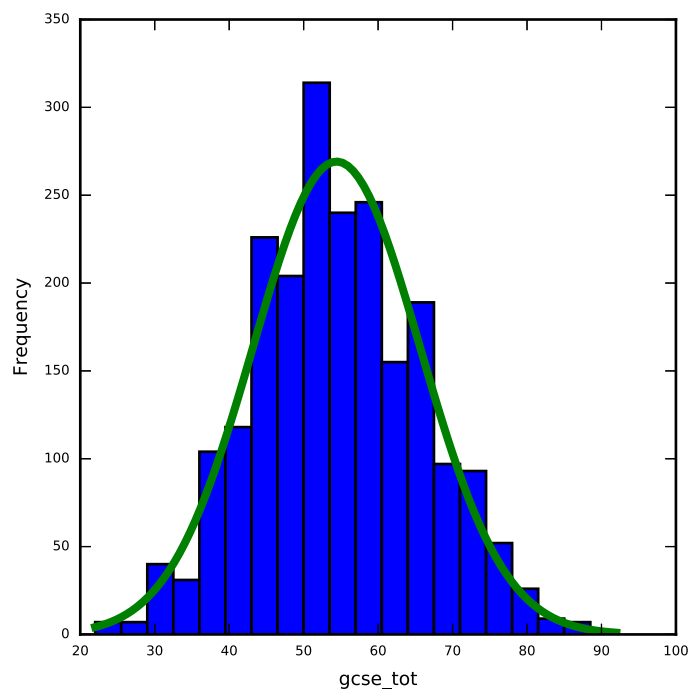


For continuous predictors we are interested in looking at summary statistics, the shape of the distribution and any unusual values. If the distribution is skewed then we might want to transform the variable before fitting it in the model although it is more important to consider transformations of the response variable and remember what is important is whether the relationship between the response and predictor is

linear. If there are unusual values we will want to check that the unusual values are correct and not errors and also whether we may want to treat the variable differently. Another possibility for unusual shaped distributions is to instead categorise the variable into ranges of values.

For predictor gcse\_tot we see the following:

<b>Name</b>	gcse_tot
<b>Observations</b>	2166
<b>Mean</b>	54.417
<b>Standard Deviation</b>	11.236
<b>Median</b>	54.0

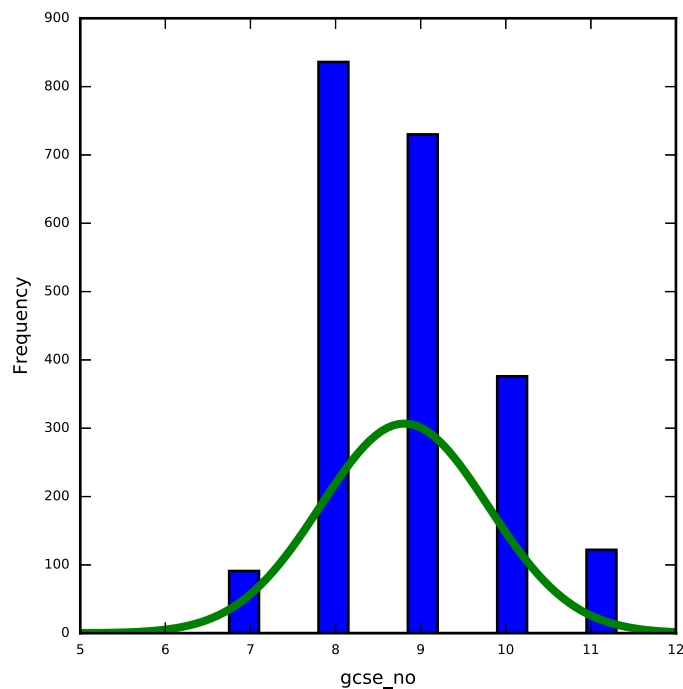


Here the median is smaller than the mean and there is significant skew to the right. The skewness value is 0.113. Here the statistical significance may be to some degree due to the large sample size as from a practical perspective values of skew less than 2 in absolute magnitude are not considered too big a skew.

There are no obvious outliers in gcse\_tot.

For predictor gcse\_no we see the following:

<b>Name</b>	gcse_no
<b>Observations</b>	2166
<b>Mean</b>	8.808
<b>Standard Deviation</b>	0.986
<b>Median</b>	9.0



The variable has only 8 unique values but is being treated as continuous.

Here the median is smaller than the mean and there is significant skew to the right. The skewness value is 0.371. Here the statistical significance may be to some degree due to the large sample size as from a practical perspective values of skew less than 2 in absolute magnitude are not considered too big a skew.

There are no obvious outliers in gcse\_no.

## Assessing the relationship between the response and individual predictors

Once we are happy with our response variable and our set of predictors we now want to have a preliminary look at them together before progressing to the univariable modelling.

For the categorical predictors it is worth looking at the mean value of the response in each category to assess if there are differences. We can then formally test this with a t-test for binary predictors or an ANOVA for predictors with more than 2 categories.

Here is a tabulation of the response, a\_point for predictor gender with category 0 having the largest mean and category 1 the smallest.

Category	N	Mean	Standard Deviation	Median
0	1223	3.526	1.749	4.0
1	943	3.509	1.766	4.0

The formal test is as follows:

There are two groups in the data:

The first group has 1223 observations with mean 3.526 standard deviation 1.75.

The second group has 943 observations with mean 3.509 standard deviation 1.767.

We are trying to test a hypothesis as to whether the two groups differ in their (population) means by a statistically significant amount. Statistical significance is related to how likely a result is to be a chance occurrence. Here we are trying to differentiate between a real difference (no matter how small) and a difference that may have occurred due to the samples we have chosen.

The mean difference is 0.017 with the first group having the larger sample mean.

We need to quantify if this difference is large relative to the variability in the data. To do this we calculate the standard error of the difference. This is a function of the variabilities in the samples from group A and group B combined with their sample sizes. The bigger the 2 variabilities the larger the standard error, whilst the smaller the variability the smaller the standard error.

For our data the standard error of the mean difference is 0.076 and we divide our observed difference by this standard error to give a test statistic with value 0.22.

This test statistic is then compared to a t distribution with degrees of freedom equal to the sum of the sample sizes in each group (2166) - 2. In this case a t distribution with 2164. This t table has values of 1.961 for  $p=0.05$  and 2.578 for  $p=0.01$ .



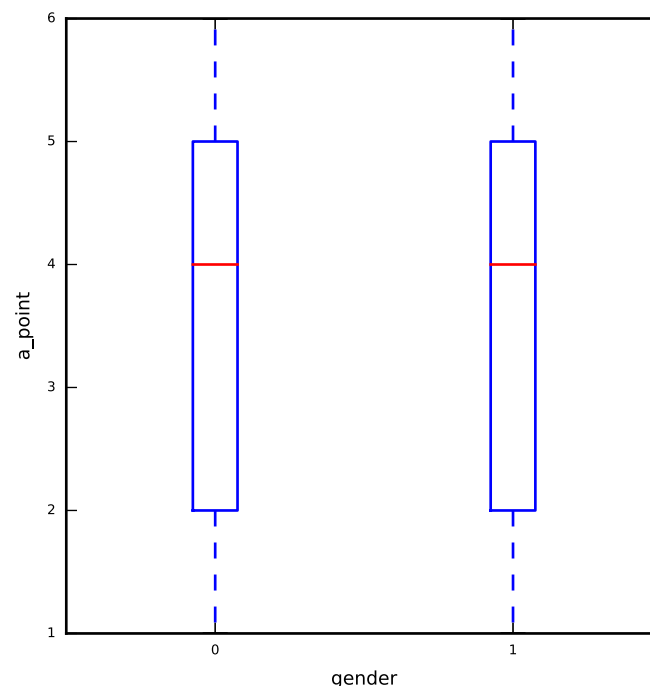
As our test statistic is  $0.22 < 1.961$  this means that the p value is  $> 0.05$  and so we cannot reject the null hypothesis.

The p-value is in fact 0.8263.

The t test assumes that the distribution of the response in each group follows a Normal distribution. We could check this by looking at histograms of the variable in each group. If we were concerned about the normality assumption then we could instead use a Mann Whitney (MW) test.

A Mann Whitney test works simply on the order (or ranks) of the responses across the two groups. So the response variable is firstly sorted and then each value is ranked. The ranks for each group are then summed and the value that is larger is compared with what would be expected if there was no difference between the groups.

In this case the MW U statistic is 579398 which for samples of size 1223 and 943 corresponds to a p value of 1.6929.

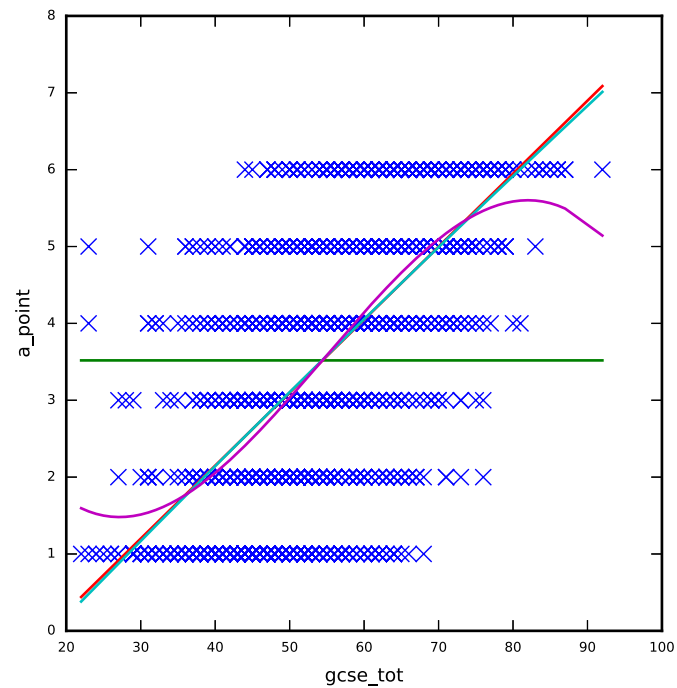


For the continuous predictors we can look at correlations with the response and scatterplots to see if there is a linear relationship.

Predictor : gcse\_tot

The Pearson correlation between a\_point and gcse\_tot is 0.607 (p value  $< 0.001$ ).

The Spearman rank correlation between a\_point and gcse\_tot is 0.614 (p value  $< 0.001$ ).

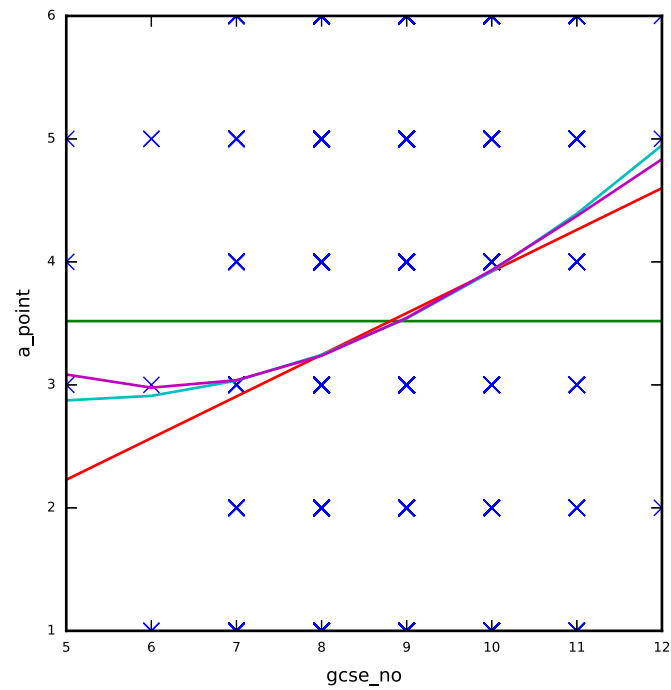


The graph includes best fitting curves for a constant, linear, quadratic and cubic relationship between a\_point and gcse\_tot. In this case a linear relationship is most appropriate.

Predictor : gcse\_no

The Pearson correlation between a\_point and gcse\_no is 0.19 (p value < 0.001).

The Spearman rank correlation between a\_point and gcse\_no is 0.183 (p value < 0.001).



The graph includes best fitting curves for a constant, linear, quadratic and cubic relationship between a\_point and gcse\_no. In this case a linear relationship is most appropriate.

## Choosing appropriate random classifications

We begin this section by deciding which of the possible random classifications to include in the modelling.

This is done by fitting combinations in turn and picking more complicated models if they make a significant improvement via a LR test. All models are displayed along with their likelihood in the table below:

Higher-level classifications	Deviance	Likelihood Ratio	p value
None	8587.81	-	-
estab	8205.24	382.57	< 0.001
estab,lea	8192.81	12.43	< 0.001

The best model based on the Likelihood has levels: estab,lea

As this is a multilevel modelling SAA we will also want to look at how the response is distributed across the levels of the model.

For this we will use the best model chosen above and look at how the variance is distributed across levels.

Variable	Coefficient	SE
Intercept	3.197	0.0876
lea Variance	0.111	0.0834
estab Variance	0.733	0.122
Level 1 Variance	2.26	0.0721

Here we see that the VPC for lea =  $0.111/3.104 = 0.0358$ , so we see that lea effects explain 3.583% of the variability in a\_point.

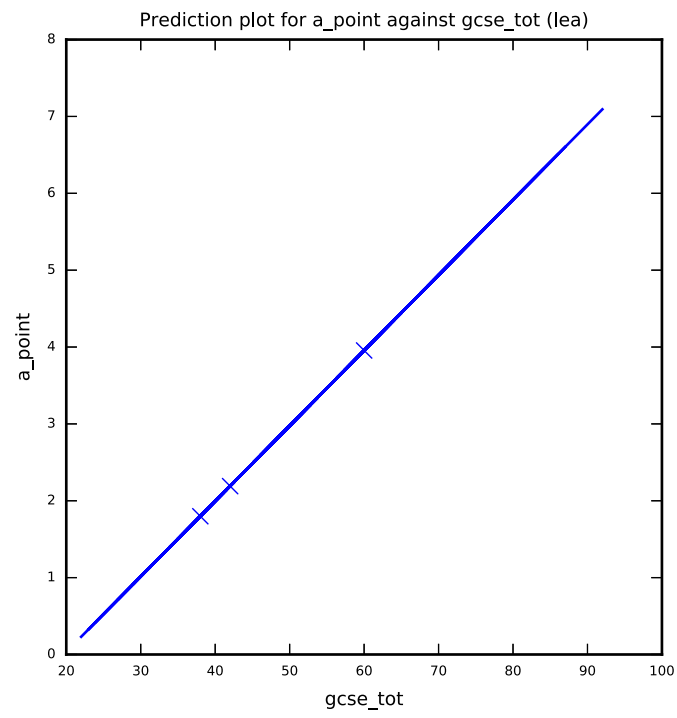
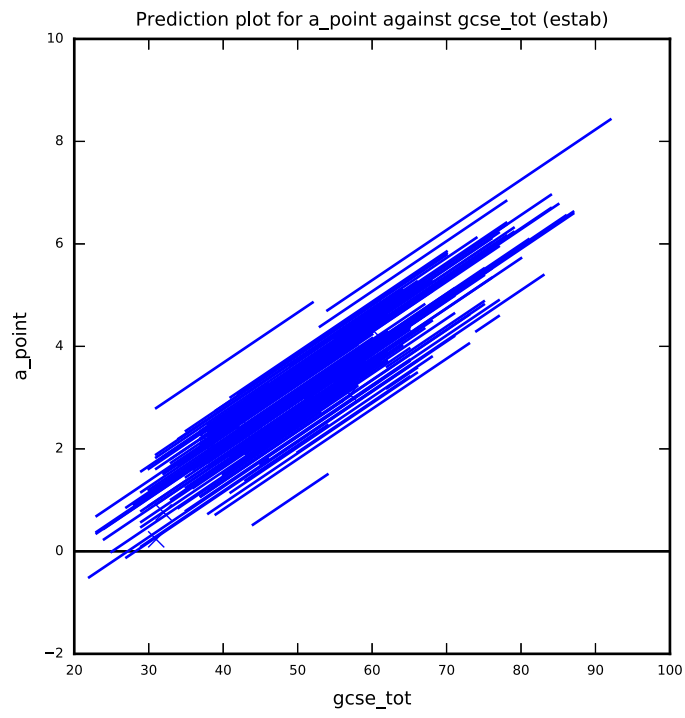
Here we see that the VPC for estab =  $0.733/3.104 = 0.236$ , so we see that estab effects explain 23.61% of the variability in a\_point.

## Performing univariable modelling

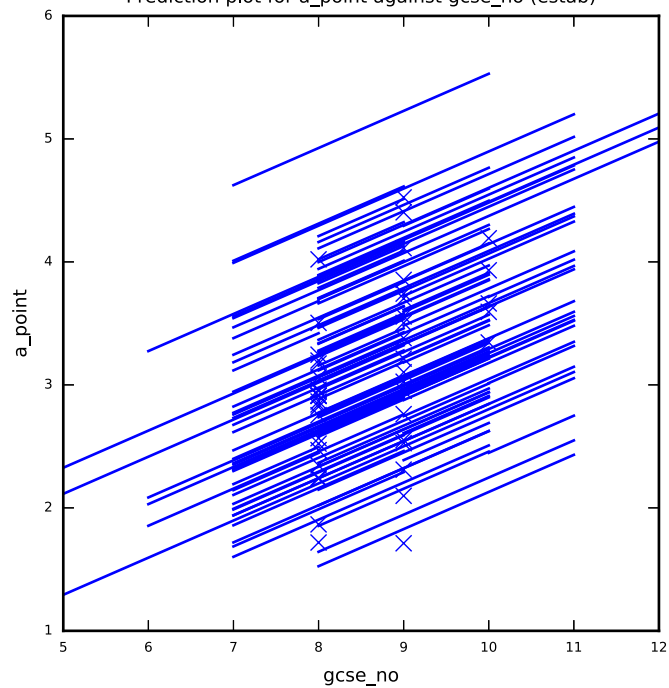
Our next step in modelling now that we have a set of potential predictors is to consider models for each predictor in turn along with a random intercept at each chosen classification from the best model in the last section. In the fixed part these models simply contain an intercept and the particular predictor and so for continuous predictors will be multilevel linear regressions and for categorical predictors will be multilevel generalisations of ANOVAs. In the table below we summarise the modelling by showing the coefficients for each predictor along with the p value comparing the model with that predictor with a Null model. This Univariable modelling step will identify a set of candidate predictors to be taken forward into the next stage of modelling.

<b>Variable</b>	<b>Coefficient</b>	<b>SE</b>	<b>p value</b>	<b>Significance</b>
<b>gcse_tot</b>	0.098	0.00297	< 0.001	***
<b>gcse_no</b>	0.302	0.0422	< 0.001	***
<b>gender_1</b>	0.108	0.0732	0.139	N/S

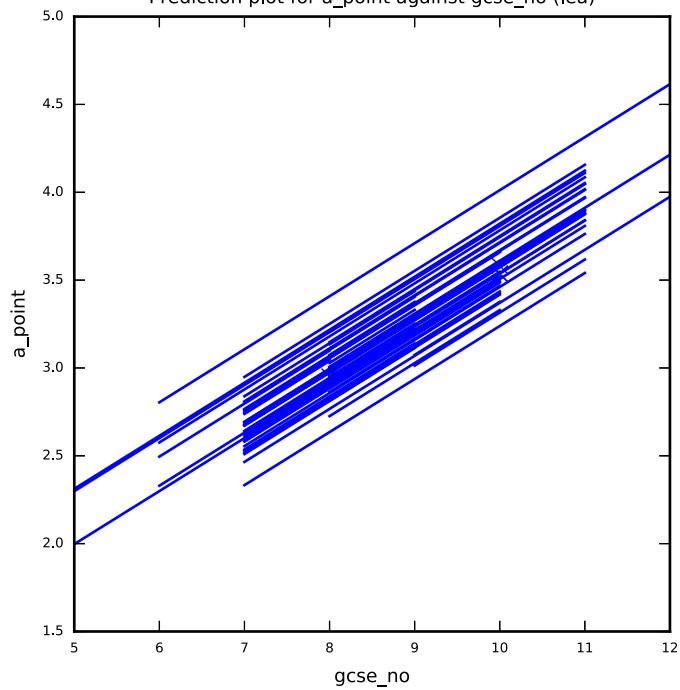
Which predictors we consider for the next stage of analysis will depend on their significance in the above table (but may in practice also depend on the size the effect and substantive interest of the variable though this is hard to automate). so the predictors to carry forward are: gcse\_no, and gcse\_tot.

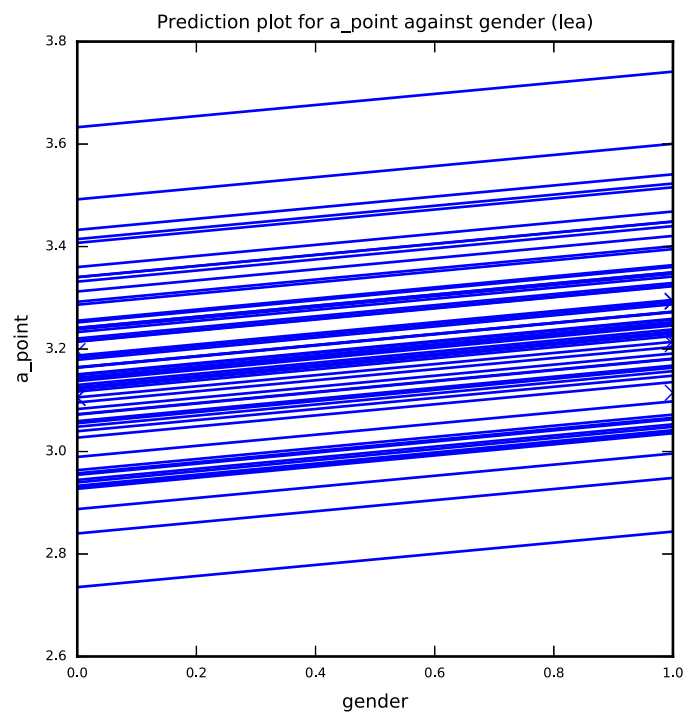
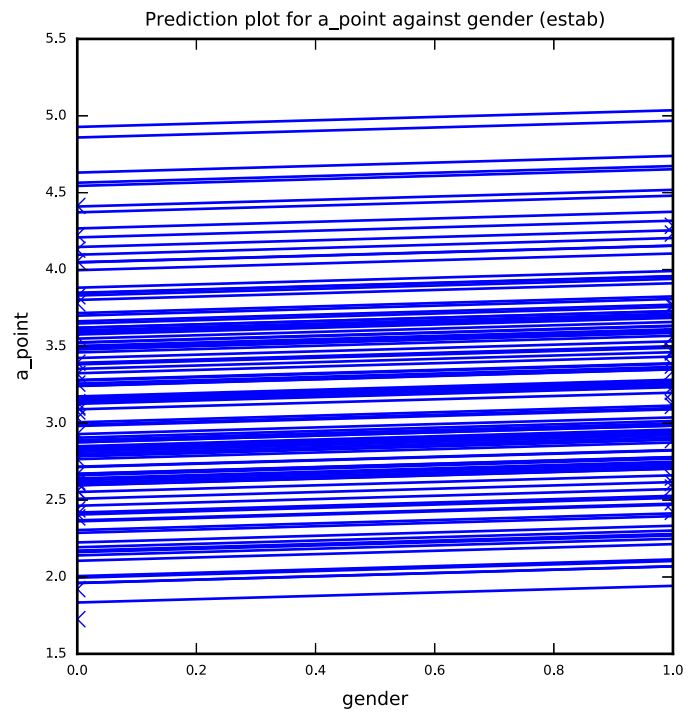


Prediction plot for a\_point against gcse\_no (estab)



Prediction plot for a\_point against gcse\_no (lea)







## Looking at correlations between predictors

Our next step is to check that none of the correlations between the predictor variables are too great as this could cause estimation problems when we add the predictors to the model together. To do this we look at all correlations between the predictor variables that have been identified as significant univariably and are thus candidates to be added to the model.

The correlations are as follows:

<b>Variables</b>	<b>Correlation</b>
<b>(gcse_no, gcse_tot)</b>	0.712
<b>(gender_1, gcse_tot)</b>	0.126
<b>(gender_1, gcse_no)</b>	0.01

Correlations greater than 0.8 (in magnitude) are worth looking at as they may result in model fitting problems when both predictors are included.

## Performing multivariable model selection - random intercept models

In this next stage we will look at the best random intercepts model using only main effects for the variables to be considered. You have chosen to perform forward pass which is a quicker method than full forward selection. It may therefore not explore as many possible models. The predictor variables are considered in turn based on their significance in the univariable analysis and each is added to the current model. If the resulting model is a significant improvement then the predictor is kept in the model otherwise it is removed. Attention then moves on to the next predictor until all predictors are considered.

You have chosen to use Likelihood ratio tests to compare models and here change in deviance will be used to indicate whether a model is better or not by comparison with an appropriate chi-squared distribution. This method is slightly slower than the alternative Wald test which we offer as for each stage a model is compared with all its submodels (with 1 predictor removed) to work out p values for each predictor.

The most significant predictor in the univariable analysis was gcse\_tot so our starting point in multivariable modelling is the model:

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SE	p value	Significance
gcse_tot	0.098	0.00297	< 0.001	*** (df=1)
Intercept	-1.926	0.166		
Between lea Variance	0.0	0.0		
Between estab Variance	0.443	0.0652		
Level 1 Variance	1.53	0.0488		

Adding variable gcse\_tot is significant and so is retained in the model.

Our next step is to consider adding variable gcse\_no to the current model.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SE	p value	Significance
gcse_tot	0.142	0.00363	< 0.001	*** (df=1)
gcse_no	-0.781	0.0422	< 0.001	*** (df=1)
Intercept	2.594	0.286		
Between lea Variance	0.0	0.0		
Between estab Variance	0.256	0.043		
Level 1 Variance	1.36	0.0433		

Adding variable gcse\_no is significant and so is retained in the model.

Our next step is to consider adding variable gender to the current model.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SE	p value	Significance
gcse_tot	0.147	0.00367	< 0.001	*** (df=1)
gcse_no	-0.823	0.0421	< 0.001	*** (df=1)
gender_1	-0.389	0.0564	< 0.001	*** (df=1)
Intercept	2.848	0.285		
Between lea Variance	0.0	0.0		
Between estab Variance	0.236	0.0407		
Level 1 Variance	1.335	0.0425		

Adding variable gender is significant and so is retained in the model.

This is our final model.

## Choosing interactions

In this section we add to the best random intercepts model with main effects found in the last section. Here we consider all possible pairwise interactions between the significant predictors already found including quadratic terms for predictors. The model selection methods used are as for the previous best random intercepts models.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SE	p value	Significance
<b>gcse_tot</b>	0.0906	0.0177	< 0.001	*** (df=1)
<b>gcse_no</b>	-0.839	0.0423	< 0.001	*** (df=1)
<b>gender_1</b>	-0.37	0.0566	< 0.001	*** (df=1)
<b>gcse_tot_X_gcse_tot</b>	0.000526	0.00016	0.001	** (df=1)
<b>Intercept</b>	4.446	0.564		
<b>Between lea Variance</b>	0.0	0.0		
<b>Between estab Variance</b>	0.236	0.0406		
<b>Level 1 Variance</b>	1.328	0.0423		

Adding variable gcse\_tot\_X\_gcse\_tot

Variable gcse\_tot\_X\_gcse\_tot significantly improved the model and so is retained in the model.

Our next step is to consider adding variable gender\_X\_gcse\_tot to the current model.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i + \beta_5 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SE	p value	Significance
<b>gcse_tot</b>	0.0898	0.0177	< 0.001	*** (df=1)
<b>gcse_no</b>	-0.829	0.0424	< 0.001	*** (df=1)
<b>gender_1</b>	-1.041	0.279	< 0.001	*** (df=1)
<b>gcse_tot_X_gcse_tot</b>	0.000486	0.000161	0.003	** (df=1)
<b>gender_1_X_gcse_tot</b>	0.0124	0.00506	0.014	* (df=1)
<b>Intercept</b>	4.519	0.564		
<b>Between lea Variance</b>	0.0	0.0		
<b>Between estab Variance</b>	0.238	0.0408		
<b>Level 1 Variance</b>	1.324	0.0421		

Adding variable gender\_X\_gcse\_tot

Variable gender\_X\_gcse\_tot significantly improved the model and so is retained in the model.

Our next step is to consider adding variable gender\_X\_gcse\_no to the current model.

$$\begin{aligned}
 a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
 & + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
 & + \beta_5 gender\_1\_X\_gcse\_no_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i
 \end{aligned}$$

Variable	Coefficient	SE	p value	Significance
gcse_tot	0.0867	0.0178	< 0.001	*** (df=1)
gcse_no	-0.791	0.0527	< 0.001	*** (df=1)
gender_1	-0.492	0.525	0.349	N/S (df=1)
gcse_tot_X_gcse_tot	0.000491	0.000161	0.002	** (df=1)
gender_1_X_gcse_tot	0.0181	0.00686	0.008	** (df=1)
gender_1_X_gcse_no	-0.098	0.0794	0.218	N/S (df=1)
Intercept	4.324	0.585		
Between lea Variance	0.0	0.0		
Between estab Variance	0.239	0.041		
Level 1 Variance	1.322	0.0421		

Adding variable gender\_X\_gcse\_no

Variable gender\_X\_gcse\_no did not significantly improve the model, so we remove it from the model and try the next predictor.

Our next step is to consider adding variable gcse\_no\_X\_gcse\_no to the current model.

$$\begin{aligned}
 a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
 & + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
 & + \beta_5 gcse\_no\_X\_gcse\_no_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i
 \end{aligned}$$

Variable	Coefficient	SE	p value	Significance
<b>gcse_tot</b>	0.0824	0.02	< 0.001	*** (df=1)
<b>gcse_no</b>	-0.499	0.423	0.239	N/S (df=1)
<b>gender_1</b>	-1.024	0.28	< 0.001	*** (df=1)
<b>gcse_tot_X_gcse_tot</b>	0.000556	0.000184	0.003	** (df=1)
<b>gender_1_X_gcse_tot</b>	0.012	0.00509	0.018	* (df=1)
<b>gcse_no_X_gcse_no</b>	-0.0186	0.0238	0.433	N/S (df=1)
<b>Intercept</b>	3.257	1.705		
<b>Between lea Variance</b>	0.0	0.0		
<b>Between estab Variance</b>	0.239	0.0409		
<b>Level 1 Variance</b>	1.323	0.0421		

Adding variable gcse\_no\_X\_gcse\_no

Variable gcse\_no\_X\_gcse\_no did not significantly improve the model, so we remove it from the model and try the next predictor.

Our next step is to consider adding variable gcse\_no\_X\_gcse\_tot to the current model.

$$\begin{aligned}
 a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
 & + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
 & + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i
 \end{aligned}$$

<b>Variable</b>	<b>Coefficient</b>	<b>SE</b>	<b>p value</b>	<b>Significance</b>
<b>gcse_tot</b>	0.135	0.0193	< 0.001	*** (df=1)
<b>gcse_no</b>	0.23	0.195	0.238	N/S (df=1)
<b>gender_1</b>	-0.886	0.279	0.002	** (df=1)
<b>gcse_tot_X_gcse_tot</b>	0.00177	0.00028	< 0.001	*** (df=1)
<b>gender_1_X_gcse_tot</b>	0.00884	0.00507	0.081	N/S (df=1)
<b>gcse_no_X_gcse_tot</b>	-0.0202	0.00363	< 0.001	*** (df=1)
<b>Intercept</b>	-1.364	1.196		
<b>Between lea Variance</b>	0.0	0.0		
<b>Between estab Variance</b>	0.239	0.0407		
<b>Level 1 Variance</b>	1.304	0.0415		

Adding variable gcse\_no\_X\_gcse\_tot

Variable gcse\_no\_X\_gcse\_tot significantly improved the model and so is retained in the model.

This is our final model.



## Adding random slopes

Having found a best model that only includes random intercepts we now investigate random slopes for significant predictor variables in the model. Here we use a simple forward pass method to look at each possible random slope in turn using the same comparison method as chosen for earlier models.

The most significant predictor in the univariable analysis was `gcse_tot` so our starting point in adding in random slopes is the model:

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{1,estab_i}^{(2)} gcse\_tot_i + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.443	1.207		
gcse_tot	0.135	0.0197		
gcse_no	0.242	0.196		
gender_1	-0.875	0.28		
gcse_tot_X_gcse_tot	0.00179	0.000283		
gender_1_X_gcse_tot	0.00862	0.00509		
gcse_no_X_gcse_tot	-0.0204	0.00364		
estab Variance(intercept)	0.464	0.369		
estab Covariance(intercept,gcse_tot)	-0.00358	0.00617		
estab Variance(gcse_tot)	5.44e-05	0.000107	0.754	N/S (df=2.0)
lea Variance(intercept)	0.0	0.0		
Level 1 Variance	1.299	0.042		

Variable gcse\_tot did not show a significant random slope, so we remove it from the random part of the model and try the next predictor.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{1,estab_i}^{(2)} gcse\_no_i + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.364	1.196		
gcse_tot	0.135	0.0193		
gcse_no	0.23	0.195		
gender_1	-0.886	0.279		
gcse_tot_X_gcse_tot	0.00177	0.00028		
gender_1_X_gcse_tot	0.00884	0.00507		
gcse_no_X_gcse_tot	-0.0202	0.00363		
estab Variance(intercept)	0.239	0.0407		
estab Covariance(intercept,gcse_no)	0.0	0.0		
estab Variance(gcse_no)	0.0	0.0	1.0	N/S (df=2.0)
lea Variance(intercept)	0.0	0.0		
Level 1 Variance	1.304	0.0415		

Variable gcse\_no did not show a significant random slope, so we remove it from the random part of the model and try the next predictor.

$$a\_point_i = \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{1,estab_i}^{(2)} gender\_1_i + u_{0,lea_i}^{(3)} + e_i$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.436	1.199		
gcse_tot	0.136	0.0193		
gcse_no	0.239	0.195		
gender_1	-0.852	0.288		
gcse_tot_X_gcse_tot	0.00178	0.00028		
gender_1_X_gcse_tot	0.00798	0.00524		
gcse_no_X_gcse_tot	-0.0203	0.00363		
estab Variance(intercept)	0.237	0.0498		
estab Covariance(intercept,gender_1)	-0.0182	0.0422		
estab Variance(gender_1)	0.0879	0.0592	0.148	N/S (df=2.0)
lea Variance(intercept)	0.0	0.0		
Level 1 Variance	1.286	0.0419		

Variable gender did not show a significant random slope, so we remove it from the random part of the model.

We have considered all predictor variables so now run our final random slopes model.

$$\begin{aligned}
 a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
 & + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
 & + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i
 \end{aligned}$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.364	1.196		
gcse_tot	0.135	0.0193		
gcse_no	0.23	0.195		
gender_1	-0.886	0.279		
gcse_tot_X_gcse_tot	0.00177	0.00028		
gender_1_X_gcse_tot	0.00884	0.00507		
gcse_no_X_gcse_tot	-0.0202	0.00363		
estab Variance(intercept)	0.239	0.0407		
lea Variance(intercept)	0.0	0.0		
Level 1 Variance	1.304	0.0415		

$$\begin{aligned}
a\_point_i &= \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
&+ \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
&+ \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + u_{1,lea_i}^{(3)} gcse\_tot_i \\
&+ e_i
\end{aligned}$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.364	1.196		
gcse_tot	0.135	0.0193		
gcse_no	0.23	0.195		
gender_1	-0.886	0.279		
gcse_tot_X_gcse_tot	0.00177	0.00028		
gender_1_X_gcse_tot	0.00884	0.00507		
gcse_no_X_gcse_tot	-0.0202	0.00363		
estab Variance(intercept)	0.239	0.0407		
lea Variance(intercept)	0.0	0.0		
lea Covariance(intercept,gcse_tot)	0.0	0.0		
lea Variance(gcse_tot)	0.0	0.0	1.0	N/S (df=2.0)
Level 1 Variance	1.304	0.0415		

Variable gcse\_tot did not show a significant random slope, so we remove it from the random part of the model and try the next predictor.

Our next step is to consider adding random slopes for the variable gcse\_no at the lea level to the current model.

$$\begin{aligned}
a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
& + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
& + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + u_{1,lea_i}^{(3)} gcse\_no_i \\
& + e_i
\end{aligned}$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.364	1.196		
gcse_tot	0.135	0.0193		
gcse_no	0.23	0.195		
gender_1	-0.886	0.279		
gcse_tot_X_gcse_tot	0.00177	0.00028		
gender_1_X_gcse_tot	0.00884	0.00507		
gcse_no_X_gcse_tot	-0.0202	0.00363		
estab Variance(intercept)	0.239	0.0407		
lea Variance(intercept)	0.0	0.0		
lea Covariance(intercept,gcse_no)	0.0	0.0		
lea Variance(gcse_no)	0.0	0.0	1.0	N/S (df=2.0)
Level 1 Variance	1.304	0.0415		

Variable gcse\_no did not show a significant random slope, so we remove it from the random part of the model and try the next predictor.

Our next step is to consider adding random slopes for the variable gender at the lea level to the current model.

$$\begin{aligned}
a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
& + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
& + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} \\
& + u_{1,lea_i}^{(3)} gender\_1_i + e_i
\end{aligned}$$

Variable	Coefficient	SD	p value	Significance
Intercept	-1.461	1.194		
gcse_tot	0.136	0.0193		
gcse_no	0.245	0.195		
gender_1	-0.835	0.287		
gcse_tot_X_gcse_tot	0.00179	0.00028		
gender_1_X_gcse_tot	0.00812	0.0052		
gcse_no_X_gcse_tot	-0.0206	0.00362		
estab Variance(intercept)	0.234	0.0413		
lea Variance(intercept)	0.0	0.0		
lea Covariance(intercept,gender_1)	0.0	0.0		
lea Variance(gender_1)	0.0543	0.0349	0.372	N/S (df=2.0)
Level 1 Variance	1.29	0.0414		

Variable gender did not show a significant random slope, so we remove it from the random part of the model.

We have considered all predictor variables so now run our final random slopes model.

$$\begin{aligned}
a\_point_i = & \beta_0 gcse\_tot_i + \beta_1 gcse\_no_i + \beta_2 gender\_1_i \\
& + \beta_3 gcse\_tot\_X\_gcse\_tot_i + \beta_4 gender\_1\_X\_gcse\_tot_i \\
& + \beta_5 gcse\_no\_X\_gcse\_tot_i + \beta_6 intercept_i + u_{0,estab_i}^{(2)} + u_{0,lea_i}^{(3)} + e_i
\end{aligned}$$

<b>Variable</b>	<b>Coefficient</b>	<b>SD</b>	<b>p value</b>	<b>Significance</b>
<b>Intercept</b>	-1.364	1.196		
<b>gcse_tot</b>	0.135	0.0193		
<b>gcse_no</b>	0.23	0.195		
<b>gender_1</b>	-0.886	0.279		
<b>gcse_tot_X_gcse_tot</b>	0.00177	0.00028		
<b>gender_1_X_gcse_tot</b>	0.00884	0.00507		
<b>gcse_no_X_gcse_tot</b>	-0.0202	0.00363		
<b>estab Variance(intercept)</b>	0.239	0.0407		
<b>lea Variance(intercept)</b>	0.0	0.0		
<b>Level 1 Variance</b>	1.304	0.0415		

This is our final random slopes model.

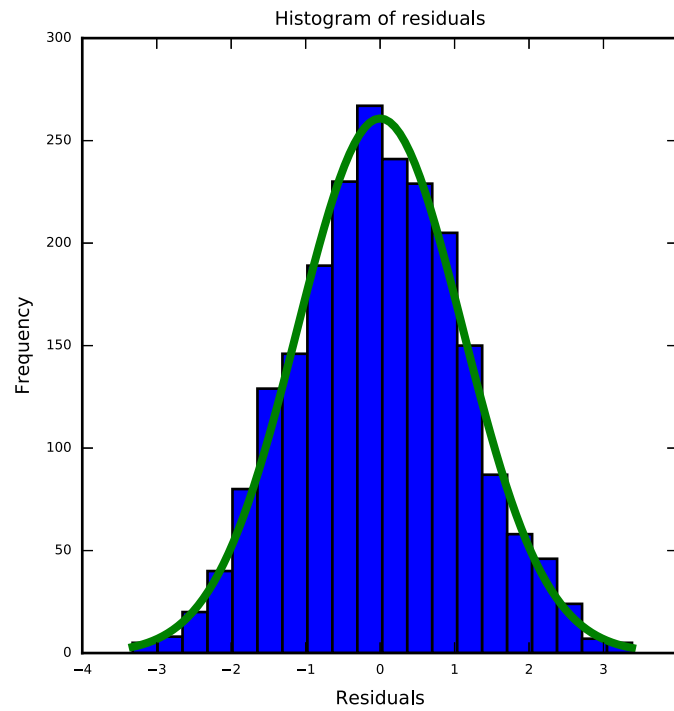




## Analysing the residuals

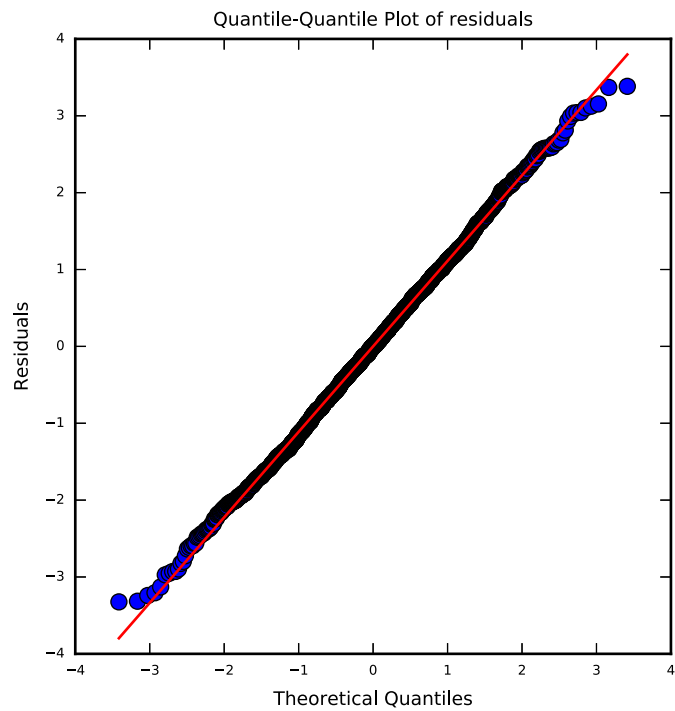
Here we look at the residuals from the model and plot them in various ways.

We start with level 1 residuals:

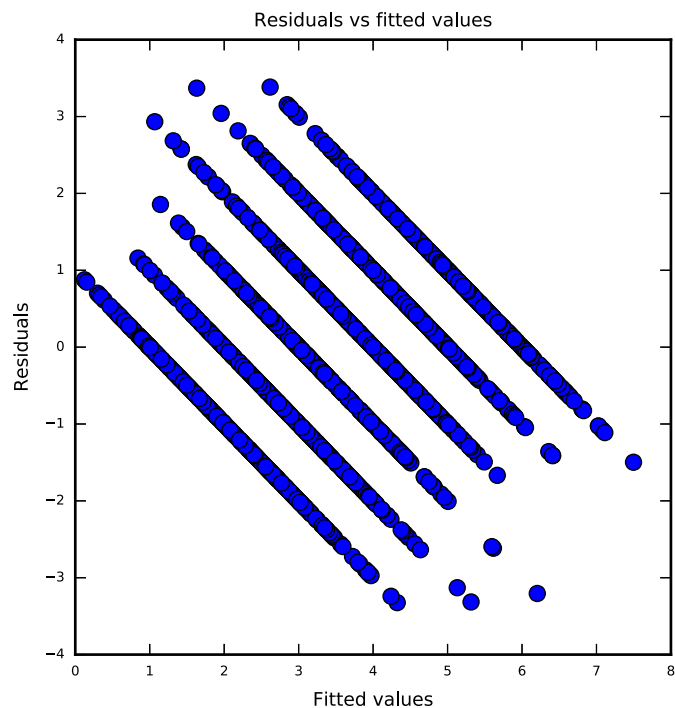


Here the distribution is reasonably symmetric with skewness value 0.037.

There are no obvious outliers in the residuals.

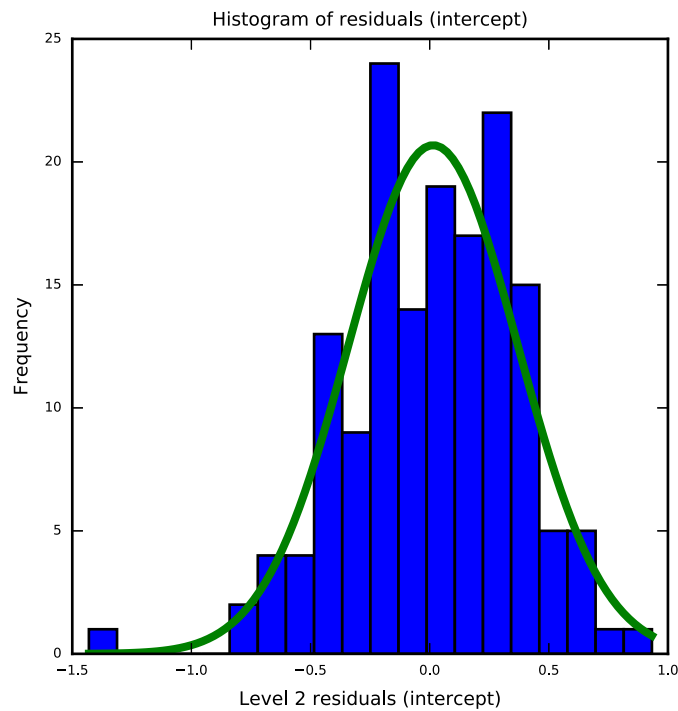


If the residuals are fairly normally distributed then the points in this graph should be close to the red line.



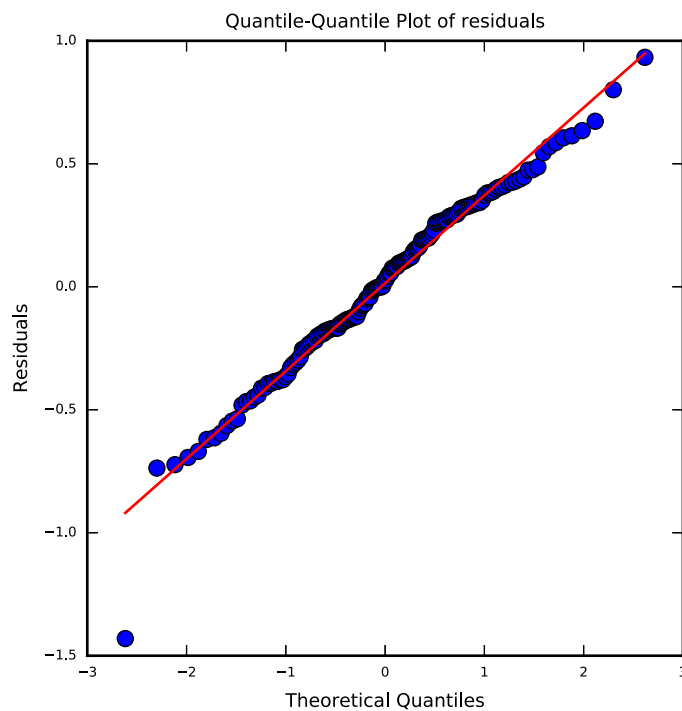
Here you should consider whether there are any patterns in this plot. Ideally we would like to see similar variability of the residuals across the range of fitted values.

Next the level 2 residuals for intercept:



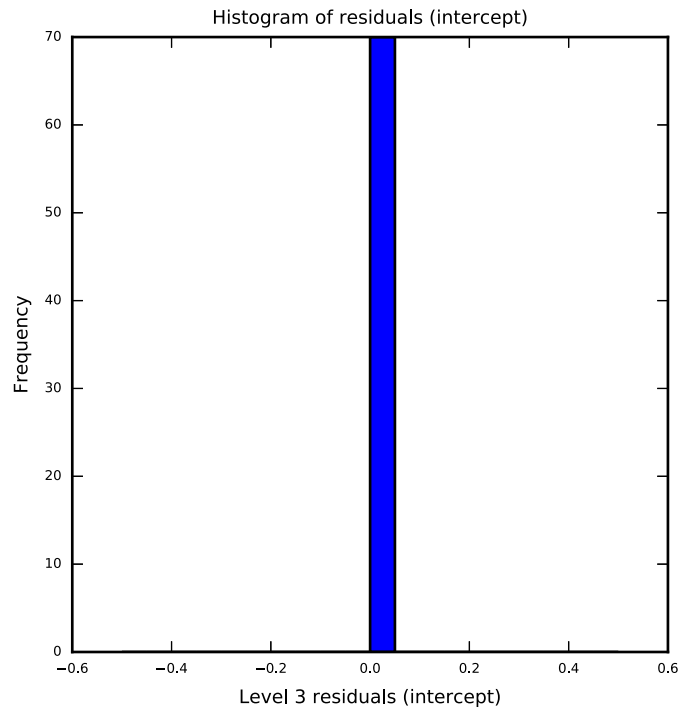
Here the median is larger than the mean and there is significant skew to the left. The skewness value is  $-0.415$ . Here the statistical significance may be to some degree due to the large sample size as from a practical perspective values of skew less than 2 in magnitude are not considered too big a skew.

There are no obvious outliers in the residuals.



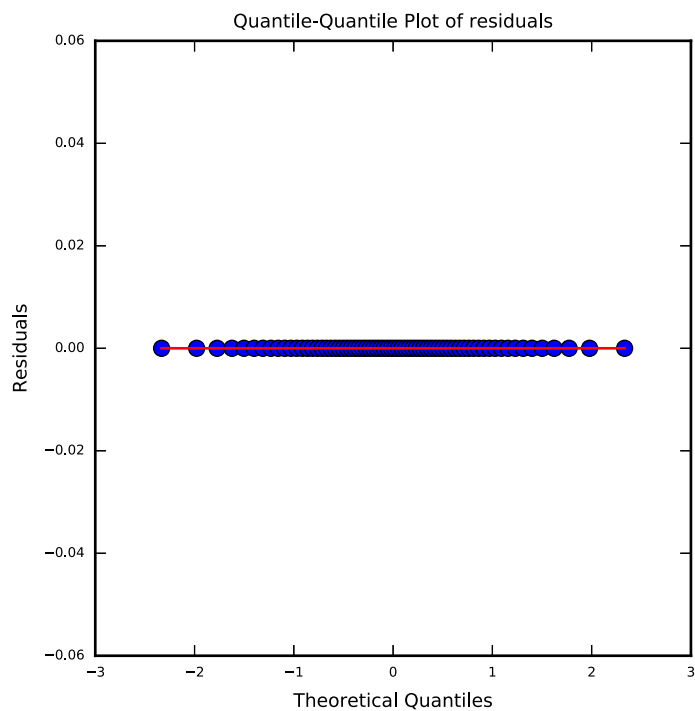
If the residuals are fairly normally distributed then the points in this graph should be close to the red line.

Next the level 3 residuals for intercept:



Here the distribution is reasonably symmetric with skewness value 0.0.

There are no obvious outliers in the residuals.



If the residuals are fairly normally distributed then the points in this graph should be close to the red line.

## Looking at predictions

Having fitted a model with several predictors we might like to represent this model graphically. This is more difficult than when we have only one predictor and so for now we consider each predictor in turn and set all other predictors to their mean values.

