# 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:** | MCMC |
| **What is the response variable:** | use |
| **What distribution are you going to assume:** | Binomial |
| **Which column contains the denominators:** | cons |
| **What link function do you wish to use:** | logit |

| | |
|---|---|
| **Which higher level classifications do you wish to consider:** | district |
| **Are there any continuous predictors that need including in all models:** | No |
| **Are there any categorical predictors that need including in all models:** | No |
| **Do you want to include any continuous predictors as candidates for inclusion in the models:** | Yes |
| **Which continuous predictors do you want to consider:** | age,d_illit,d_pray |
| **Do you want to include any categorical predictors as candidates for inclusion in the models:** | Yes |
| **Which categorical predictors do you want to consider:** | lc,urban,educ,hindu |
| **What selection type do you require:** | Forward pass |
| **Do you want to test for random slopes:** | No |
| **Do you want to test for interactions:** | Yes |
| **How do you wish to compare models:** | Wald |
| **How long to burnin for:** | 500 |

| | |
|---|---|
| **How long to then run chains for:** | 2000 |
| **What is the minimum ESS at which to stop (use 0 to just run for number last input):** | 200 |
| **Do you want to use orthogonal parameterisation:** | Yes |
| **What change in DIC denotes a better model:** | 1 |

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: use, cons, district, age, d_illit, d_pray, lc, urban, educ, hindu. There are 0 (0.0%) rows that get deleted This results in a dataset of 1934 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.

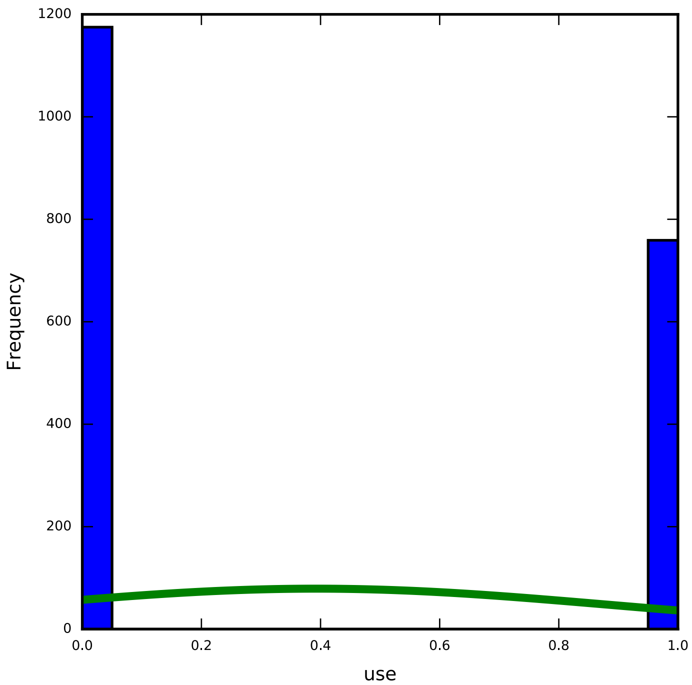
## Exploring the response

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

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

| | |
|---|---|
| **Observations** | 1934 |
| **Mean** | 0.392 |
| **Standard Deviation** | 0.488 |
| **Median** | 0.0 |

We also look at a histogram of use to see what it looks like - noting that for a Binomial model this is of less interest as it will simply look like a bar graph.




Here the median is smaller than the mean and there is significant skew to the right. The skewness value is 0.441. 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 are not considered too big a skew.

There are no obvious outliers in use.

# Exploring the predictors individually

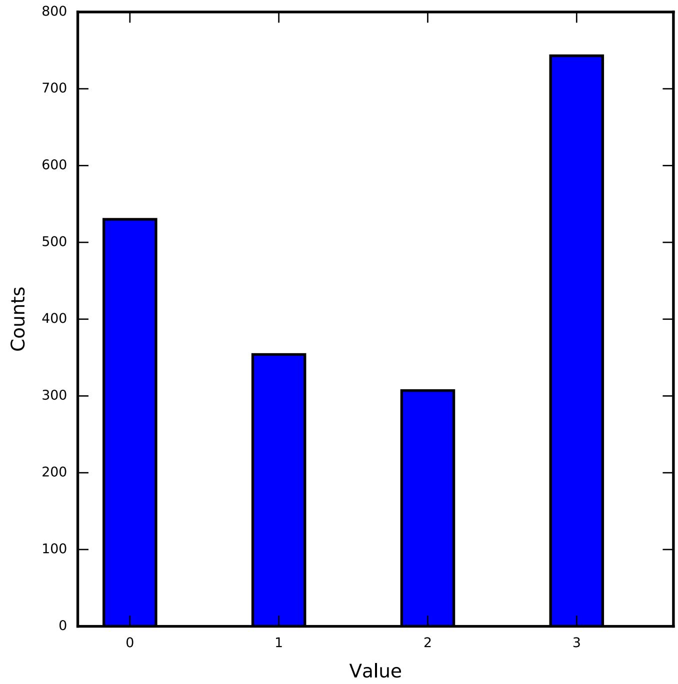
We can also look at each of the predictor variables in turn in isolation.

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 lc we see the following:

| lc | N | Percentage |
|---|---|---|
| **0** | 530 | 27.404 |
| **1** | 354 | 18.304 |
| **2** | 307 | 15.874 |
| **3** | 743 | 38.418 |
| **Total** | 1934 | 100 |

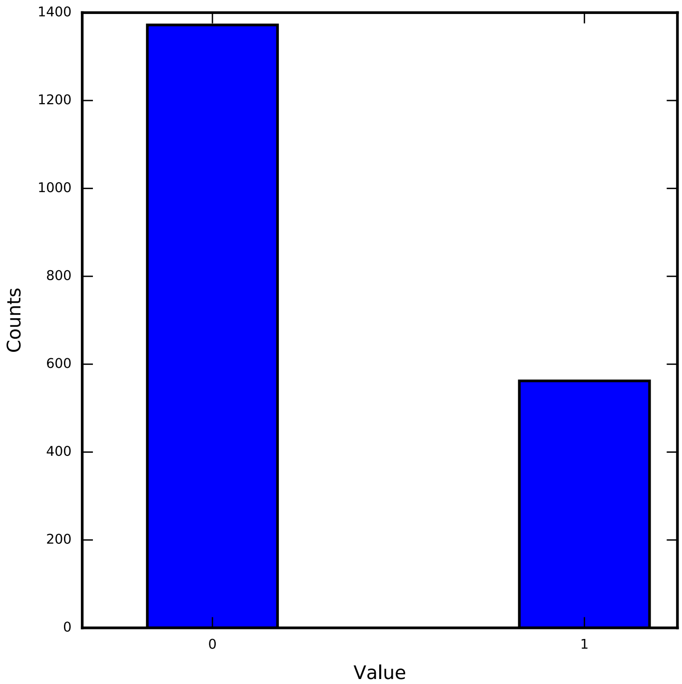
None of the categories of lc have fewer than 5 observations.




For predictor urban we see the following:

| urban | N | Percentage |
|---|---|---|
| **0** | 1372 | 70.941 |
| **1** | 562 | 29.059 |
| **Total** | 1934 | 100 |

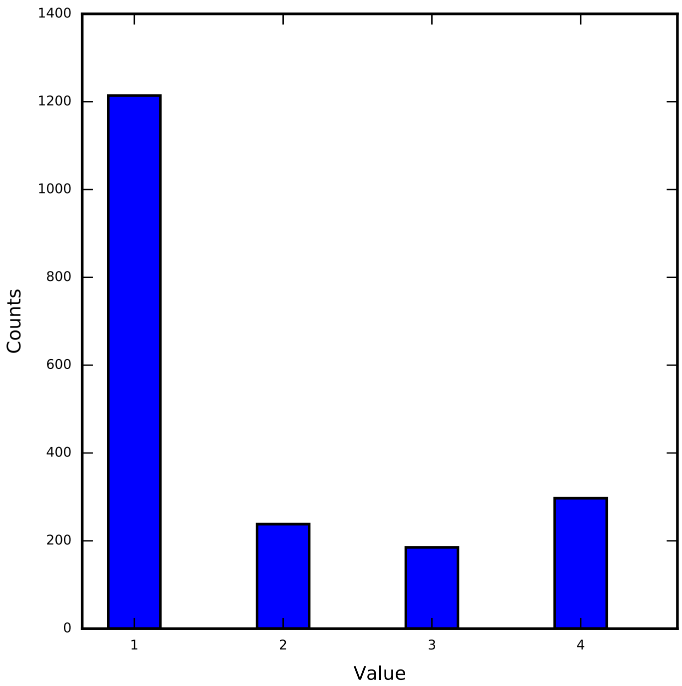
None of the categories of urban have fewer than 5 observations.




For predictor educ we see the following:

| educ | N | Percentage |
|---|---|---|
| **1** | 1214 | 62.771 |
| **2** | 238 | 12.306 |
| **3** | 185 | 9.566 |
| **4** | 297 | 15.357 |
| **Total** | 1934 | 100 |

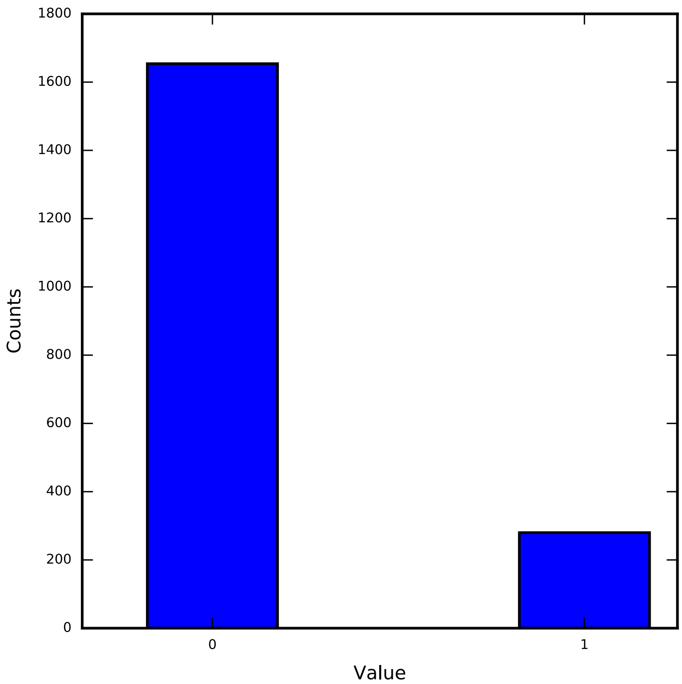
None of the categories of educ have fewer than 5 observations.




For predictor hindu we see the following:

| hindu | N | Percentage |
|---:|:---:|:---:|
| **0** | 1654 | 85.522 |
| **1** | 280 | 14.478 |
| **Total** | 1934 | 100 |

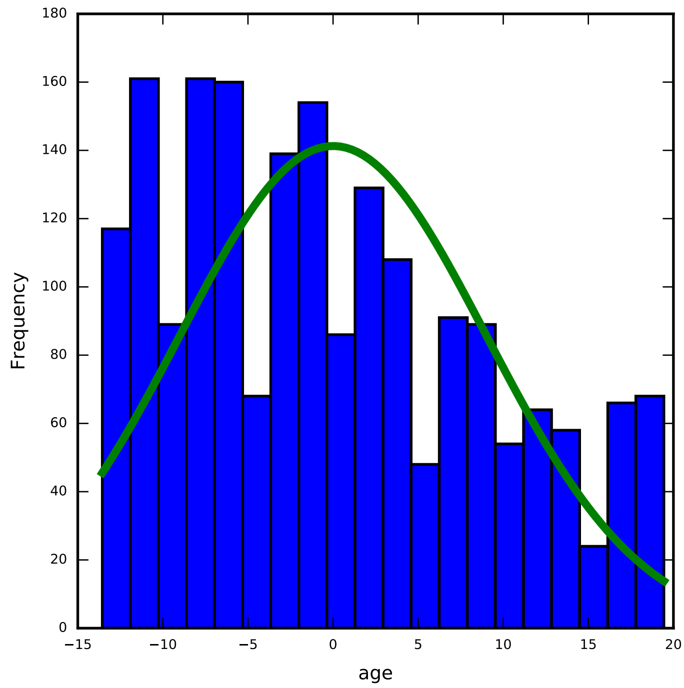
None of the categories of hindu have fewer than 5 observations.




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 age we see the following:

| | |
|---|---|
| **Name** | age |
| **Observations** | 1934 |
| **Mean** | 0.002 |
| **Standard Deviation** | 9.011 |
| **Median** | -1.56 |

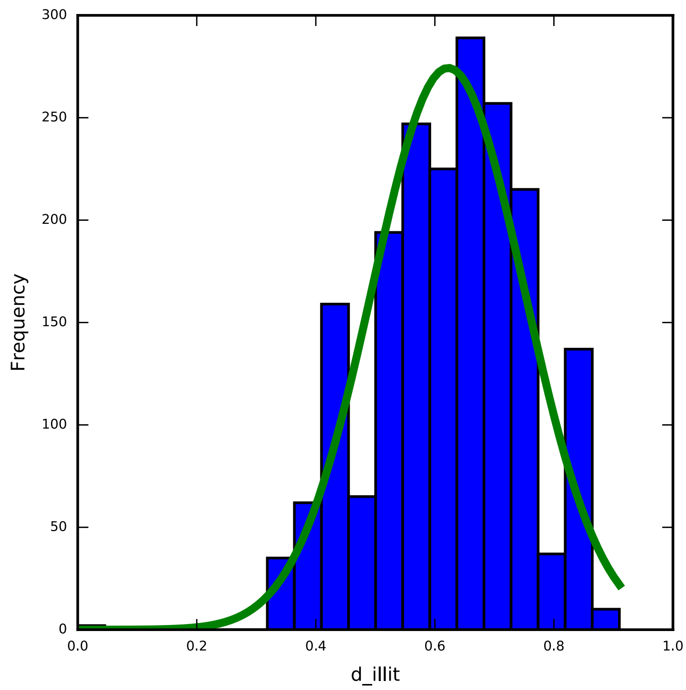



Here the median is smaller than the mean and there is significant skew to the right. The skewness value is 0.441. Here the statistical significance may be to some degree due to the large sample size as from a practical perspective values of skew less than 2 in absolute magnitude are not considered too big a skew.

There are no obvious outliers in age.

For predictor d_illit we see the following:

| | |
|---|---|
| **Name** | d_illit |
| **Observations** | 1934 |
| **Mean** | 0.622 |
| **Standard Deviation** | 0.128 |
| **Median** | 0.63 |

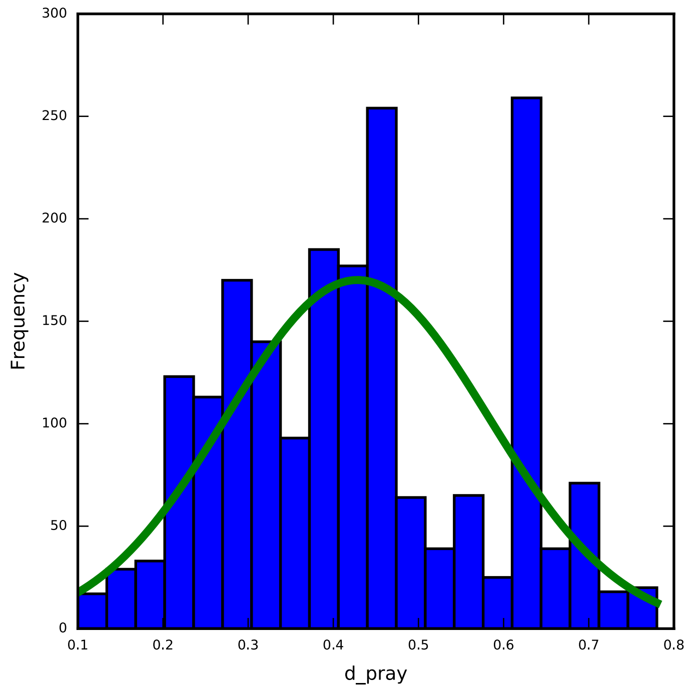



Here the median is larger than the mean and there is significant skew to the left. The skewness value is -0.323. Here the statistical significance may be to some degree due to the large sample size as from a practical perspective values of skew less than 2 in absolute magnitude are not considered too big a skew.

There are no obvious outliers in d_illit.

For predictor d_pray we see the following:

| | |
|---|---|
| **Name** | d_pray |
| **Observations** | 1934 |
| **Mean** | 0.428 |
| **Standard Deviation** | 0.154 |
| **Median** | 0.43 |




Here the median is smaller than the mean and there is significant skew to the right. The skewness value is 0.251. 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 d_pray.

## 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 tabulating the response for each category to look at whether patterns differ. We can formally test this with a chi-squared test.

We will investigate categorical variable lc. To do a chi-squared test we start by tabulated observed counts and totals:

| Observed | use=0 | use=1 | Total |
|---:|:---:|:---:|:---:|
| **lc=0** | 397 | 133 | 530 |
| **lc=1** | 190 | 164 | 354 |
| **lc=2** | 160 | 147 | 307 |
| **lc=3** | 428 | 315 | 743 |
| **Total** | 1175 | 759 | 1934 |

We can therefore work out the expected counts from the margins of the observed data.

And so we expect

E(use = 0, lc = 0) = Total use = 0 * Total lc = 0 / grand total = 1175 * 530 / 1934 = 322.0.
E(use = 1, lc = 0) = Total use = 1 * Total lc = 0 / grand total = 759 * 530 / 1934 = 208.0.
E(use = 0, lc = 1) = Total use = 0 * Total lc = 1 / grand total = 1175 * 354 / 1934 = 215.07.
E(use = 1, lc = 1) = Total use = 1 * Total lc = 1 / grand total = 759 * 354 / 1934 = 138.93.
E(use = 0, lc = 2) = Total use = 0 * Total lc = 2 / grand total = 1175 * 307 / 1934 = 186.52.
E(use = 1, lc = 2) = Total use = 1 * Total lc = 2 / grand total = 759 * 307 / 1934 = 120.48.
E(use = 0, lc = 3) = Total use = 0 * Total lc = 3 / grand total = 1175 * 743 / 1934 =

451.41.
E(use = 1, lc = 3) = Total use = 1 * Total lc = 3 / grand total = 759 * 743 / 1934 = 291.59.

So the table of expected counts is:

| Expected | use=0 | use=1 | Total |
|---|---|---|---|
| **lc=0** | 322.0 | 208.0 | 530.0 |
| **lc=1** | 215.07 | 138.93 | 354.0 |
| **lc=2** | 186.52 | 120.48 | 307.0 |
| **lc=3** | 451.41 | 291.59 | 743.0 |
| **Total** | 1175.0 | 759.0 | 1934.0 |

We next look at differences between what we observe and expect in each cell. We square these values so that every difference is positive and scale by the expected counts so that more frequently expected cells aren't overly influential. So for example for use=0, lc=0 (O-E)^2/E = (397-322.0)^2/322.0=17.47. This statistic is shown in tabular form below:

| (O-E)^2/E | use=0 | use=1 |
|---|---|---|
| **lc=0** | 17.47 | 27.04 |
| **lc=1** | 2.92 | 4.52 |
| **lc=2** | 3.77 | 5.84 |
| **lc=3** | 1.21 | 1.88 |

The test statistic for a chi-squared test is found by summing the values of this table so:

Chisq=17.47+27.04+2.92+4.52+3.77+5.84+1.21+1.88=64.66.

This is compared with a chi-squared table with degrees of freedom = (number of columns -1)x(number of rows - 1) =

(4-1)x(2-1)=3.

Looking up the chi-squared table the value for p=0.05 is 7.81 and for p=0.01 = 11.34

As 64.66 > 11.34 our p value is less than 0.01 and we have strong evidence to reject the null hypothesis (at the p=0.01) level.

The p-value is in fact less than 0.0001.

We will investigate categorical variable urban. To do a chi-squared test we start by tabulated observed counts and totals:

| Observed | use=0 | use=1 | Total |
|---|---|---|---|
| **urban=0** | 903 | 469 | 1372 |
| **urban=1** | 272 | 290 | 562 |
| **Total** | 1175 | 759 | 1934 |

We can therefore work out the expected counts from the margins of the observed data.

And so we expect

E(use = 0, urban = 0) = Total use = 0 * Total urban = 0 / grand total = 1175 * 1372 / 1934 = 833.56.
E(use = 1, urban = 0) = Total use = 1 * Total urban = 0 / grand total = 759 * 1372 / 1934 = 538.44.
E(use = 0, urban = 1) = Total use = 0 * Total urban = 1 / grand total = 1175 * 562 / 1934 = 341.44.
E(use = 1, urban = 1) = Total use = 1 * Total urban = 1 / grand total = 759 * 562 / 1934 = 220.56.

So the table of expected counts is:

| Expected | use=0 | use=1 | Total |
|---|---|---|---|
| **urban=0** | 833.56 | 538.44 | 1372.0 |
| **urban=1** | 341.44 | 220.56 | 562.0 |
| **Total** | 1175.0 | 759.0 | 1934.0 |

We next look at differences between what we observe and expect in each cell. We square these values so that every difference is positive and scale by the expected counts so that more frequently expected cells aren't overly influential. So for example for use=0, urban=0 (O-E)^2/E = (903-833.56)^2/833.56=5.79. This statistic is shown in tabular form below:

| (O-E)^2/E | use=0 | use=1 |
|---:|:---:|:---:|
| **urban=0** | 5.79 | 8.96 |
| **urban=1** | 14.12 | 21.86 |

The test statistic for a chi-squared test is found by summing the values of this table so:

Chisq=5.79+8.96+14.12+21.86=50.73.

This is compared with a chi-squared table with degrees of freedom = (number of columns -1)x(number of rows - 1) =

(2-1)x(2-1)=1.

Looking up the chi-squared table the value for p=0.05 is 3.84 and for p=0.01 = 6.63

As 50.73 > 6.63 our p value is less than 0.01 and we have strong evidence to reject the null hypothesis (at the p=0.01) level.

The p-value is in fact less than 0.0001.

We will investigate categorical variable educ. To do a chi-squared test we start by tabulated observed counts and totals:

| Observed | use=0 | use=1 | Total |
|---:|:---:|:---:|:---:|
| **educ=1** | 837 | 377 | 1214 |
| **educ=2** | 137 | 101 | 238 |
| **educ=3** | 93 | 92 | 185 |
| **educ=4** | 108 | 189 | 297 |
| **Total** | 1175 | 759 | 1934 |

We can therefore work out the expected counts from the margins of the observed data.

And so we expect

E(use = 0, educ = 1) = Total use = 0 * Total educ = 1 / grand total = 1175 * 1214 / 1934 = 737.56.
E(use = 1, educ = 1) = Total use = 1 * Total educ = 1 / grand total = 759 * 1214 / 1934 = 476.44.

E(use = 0, educ = 2) = Total use = 0 * Total educ = 2 / grand total = 1175 * 238 / 1934 = 144.6.
E(use = 1, educ = 2) = Total use = 1 * Total educ = 2 / grand total = 759 * 238 / 1934 = 93.4.
E(use = 0, educ = 3) = Total use = 0 * Total educ = 3 / grand total = 1175 * 185 / 1934 = 112.4.
E(use = 1, educ = 3) = Total use = 1 * Total educ = 3 / grand total = 759 * 185 / 1934 = 72.6.
E(use = 0, educ = 4) = Total use = 0 * Total educ = 4 / grand total = 1175 * 297 / 1934 = 180.44.
E(use = 1, educ = 4) = Total use = 1 * Total educ = 4 / grand total = 759 * 297 / 1934 = 116.56.

So the table of expected counts is:

| Expected | use=0 | use=1 | Total |
|---|---|---|---|
| **educ=1** | 737.56 | 476.44 | 1214.0 |
| **educ=2** | 144.6 | 93.4 | 238.0 |
| **educ=3** | 112.4 | 72.6 | 185.0 |
| **educ=4** | 180.44 | 116.56 | 297.0 |
| **Total** | 1175.0 | 759.0 | 1934.0 |

We next look at differences between what we observe and expect in each cell. We square these values so that every difference is positive and scale by the expected counts so that more frequently expected cells aren't overly influential. So for example for use=0, educ=1 (O-E)^2/E = (837-737.56)^2/737.56=13.41. This statistic is shown in tabular form below:

| (O-E)^2/E | use=0 | use=1 |
|---|---|---|
| **educ=1** | 13.41 | 20.75 |
| **educ=2** | 0.4 | 0.62 |
| **educ=3** | 3.35 | 5.18 |
| **educ=4** | 29.08 | 45.02 |

The test statistic for a chi-squared test is found by summing the values of this table so:

Chisq=13.41+20.75+0.4+0.62+3.35+5.18+29.08+45.02=117.81.

This is compared with a chi-squared table with degrees of freedom = (number of columns -1)x(number of rows - 1) =

(4-1)x(2-1)=3.

Looking up the chi-squared table the value for p=0.05 is 7.81 and for p=0.01 = 11.34

As 117.81 > 11.34 our p value is less than 0.01 and we have strong evidence to reject the null hypothesis (at the p=0.01) level.

The p-value is in fact less than 0.0001.

We will investigate categorical variable hindu. To do a chi-squared test we start by tabulated observed counts and totals:

| Observed | use=0 | use=1 | Total |
|---:|:---:|:---:|:---:|
| **hindu=0** | 1017 | 637 | 1654 |
| **hindu=1** | 158 | 122 | 280 |
| **Total** | 1175 | 759 | 1934 |

We can therefore work out the expected counts from the margins of the observed data.

And so we expect

E(use = 0, hindu = 0) = Total use = 0 * Total hindu = 0 / grand total = 1175 * 1654 / 1934 = 1004.89.
E(use = 1, hindu = 0) = Total use = 1 * Total hindu = 0 / grand total = 759 * 1654 / 1934 = 649.11.
E(use = 0, hindu = 1) = Total use = 0 * Total hindu = 1 / grand total = 1175 * 280 / 1934 = 170.11.
E(use = 1, hindu = 1) = Total use = 1 * Total hindu = 1 / grand total = 759 * 280 / 1934 = 109.89.

So the table of expected counts is:

| Expected | use=0 | use=1 | Total |
|---|---|---|---|
| **hindu=0** | 1004.89 | 649.11 | 1654.0 |
| **hindu=1** | 170.11 | 109.89 | 280.0 |
| **Total** | 1175.0 | 759.0 | 1934.0 |

We next look at differences between what we observe and expect in each cell. We square these values so that every difference is positive and scale by the expected counts so that more frequently expected cells aren’t overly influential. So for example for use=0, hindu=0 (O-E)^2/E = (1017-1004.89)^2/1004.89=0.15. This statistic is shown in tabular form below:

| (O-E)^2/E | use=0 | use=1 |
|---|---|---|
| **hindu=0** | 0.15 | 0.23 |
| **hindu=1** | 0.86 | 1.34 |

The test statistic for a chi-squared test is found by summing the values of this table so:

Chisq=0.15+0.23+0.86+1.34=2.57.

This is compared with a chi-squared table with degrees of freedom = (number of columns -1)x(number of rows - 1) =

(2-1)x(2-1)=1.

Looking up the chi-squared table the value for p=0.05 is 3.84 and for p=0.01 = 6.63

As our test statistic is 2.57 %lt; 3.84 this means that the p value is > 0.05 and so we cannot reject the null hypothesis.

The p-value is in fact 0.1089.

For the continuous predictors it is worth looking at the mean value of each predictor for the 0 and 1 responses to assess if there is any difference. We can formally test this with a t-test.

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

| Category | N | Mean | Standard Deviation | Median |
|---|---|---|---|---|
| **0** | 1175 | -0.208 | 9.707 | -1.56 |
| **1** | 759 | 0.327 | 7.802 | -0.56 |

The formal test is as follows:

There are two groups in the data:
The first group has 1175 observations with mean -0.208 standard deviation 9.711.
The second group has 759 observations with mean 0.327 standard deviation 7.807.
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 occurance. 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 0.534 with the second 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 0.401 and we divide our observed difference by this standard error to give a test statistic with value 1.334.
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 (1934) - 2. In this case a t distribution with 1932. This t table has values of 1.961 for p=0.05 and 2.578 for p=0.01.

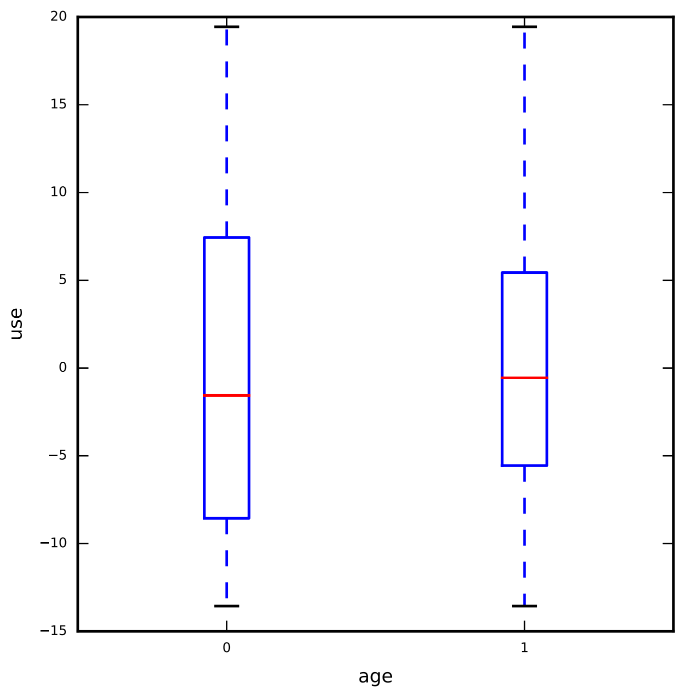
As our test statistic is 1.334 < 1.961 this means that the p value is > 0.05 and so we cannot reject the null hypothesis.

The p-value is in fact 0.1825. .

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 413204 which for samples of size 1175 and 759 corresponds to a p value of 0.0127.




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

| Category | N | Mean | Standard Deviation | Median |
|---|---|---|---|---|
| **0** | 1175 | 0.639 | 0.124 | 0.65 |
| **1** | 759 | 0.597 | 0.13 | 0.62 |

The formal test is as follows:

There are two groups in the data:
The first group has 1175 observations with mean 0.639 standard deviation 0.124.
The second group has 759 observations with mean 0.597 standard deviation 0.13.
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 occurance. 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 0.042 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 0.006 and we divide our observed difference by this standard error to give a test statistic with value 7.077.
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 (1934) - 2. In this case a t distribution with 1932. This t table has values of 1.961 for p=0.05 and 2.578 for p=0.01.

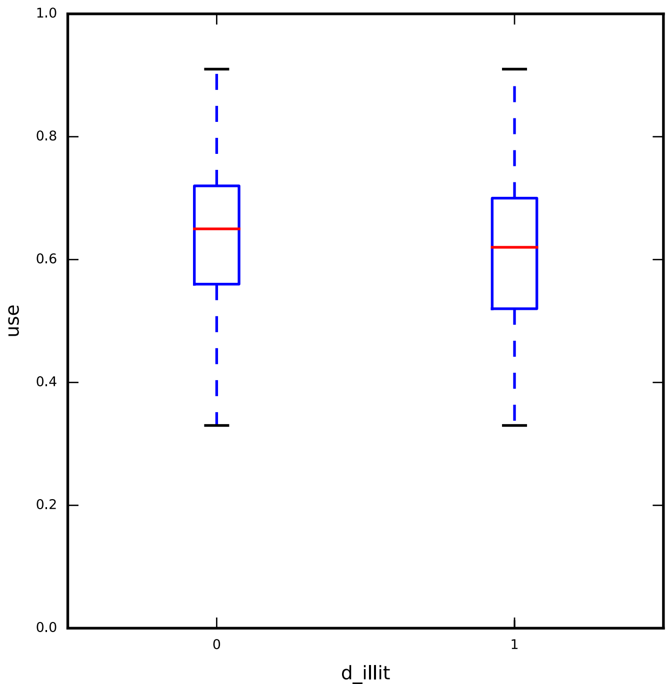
As 7.077 > 2.578 our p value is less than 0.01 and we have strong evidence to reject the null hypothesis (at the p=0.01 level).

The p-value is in fact less than 0.0001..

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 521432 which for samples of size 1175 and 759 corresponds to a p value of less than 0.0001.




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

| Category | N | Mean | Standard Deviation | Median |
|---|---|---|---|---|
| **0** | 1175 | 0.436 | 0.157 | 0.43 |
| **1** | 759 | 0.417 | 0.149 | 0.43 |

The formal test is as follows:

There are two groups in the data:
The first group has 1175 observations with mean 0.436 standard deviation 0.157.
The second group has 759 observations with mean 0.417 standard deviation 0.149.
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 occurance. 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 0.018 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 0.007 and we divide our observed difference by this standard error to give a test statistic with value 2.603. 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 (1934) - 2. In this case a t distribution with 1932. This t table has values of 1.961 for p=0.05 and 2.578 for p=0.01.

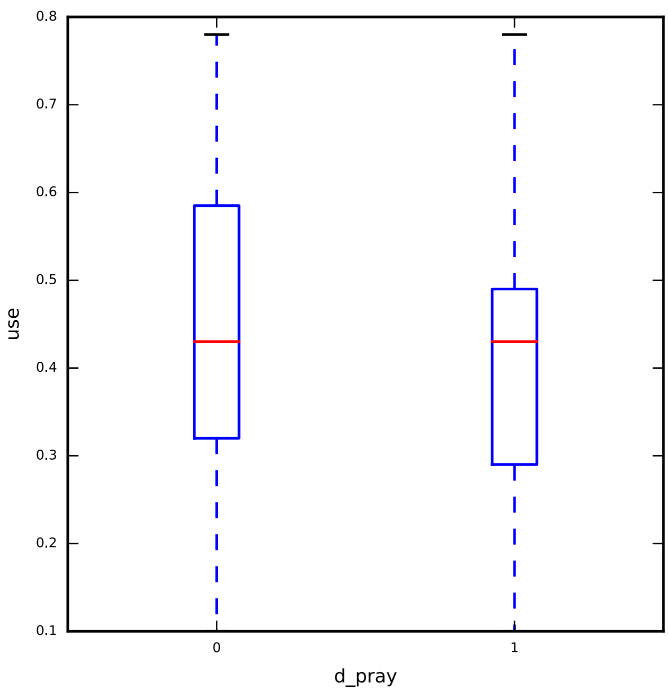
As 2.603 > 2.578 our p value is less than 0.01 and we have strong evidence to reject the null hypothesis (at the p=0.01 level).

The p-value is in fact 0.0093. .

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 473639 which for samples of size 1175 and 759 corresponds to a p value of 0.04131.



## 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 all possible combinations and picking the model with the lowest DIC. All models are displayed along with their DIC values in the table below:

| Higher-level classifications | DIC |
|---:|:---:|
| **None** | 2592.77 |
| **district** | 2515.04 |

The best model based on the DIC has classifications: district

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 | ESS |
|---:|:---:|:---:|:---:|
| **Intercept** | -0.535 | 0.0832 | 203 |
| **district Variance** | 0.269 | 0.0891 | 386 |

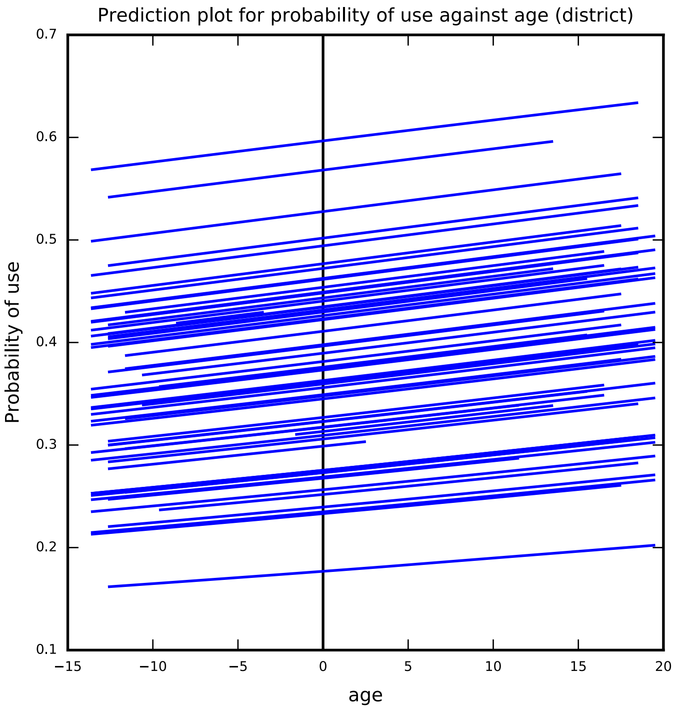
Here we see that the VPC for district = 0.269/3.559 = 0.0756, so we see that district effects explain 7.565% of the variability in use.

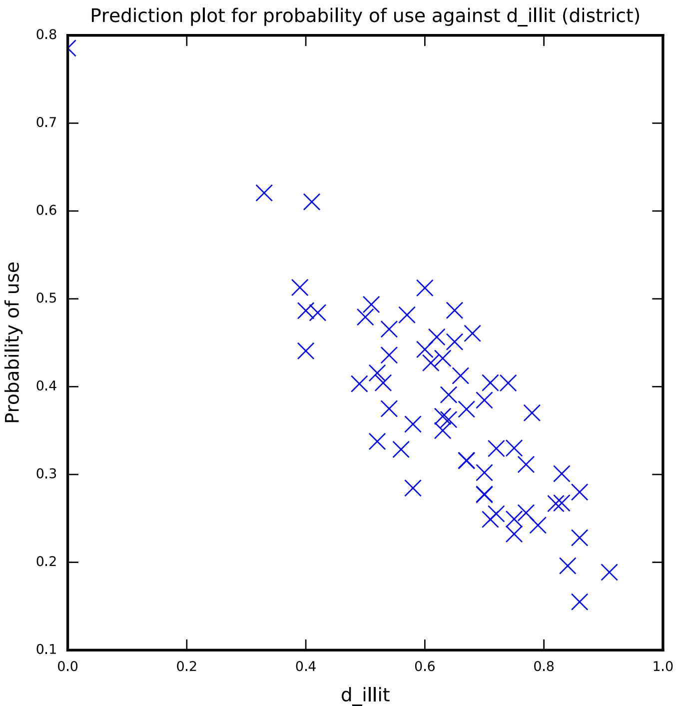
## Performing univariable modelling

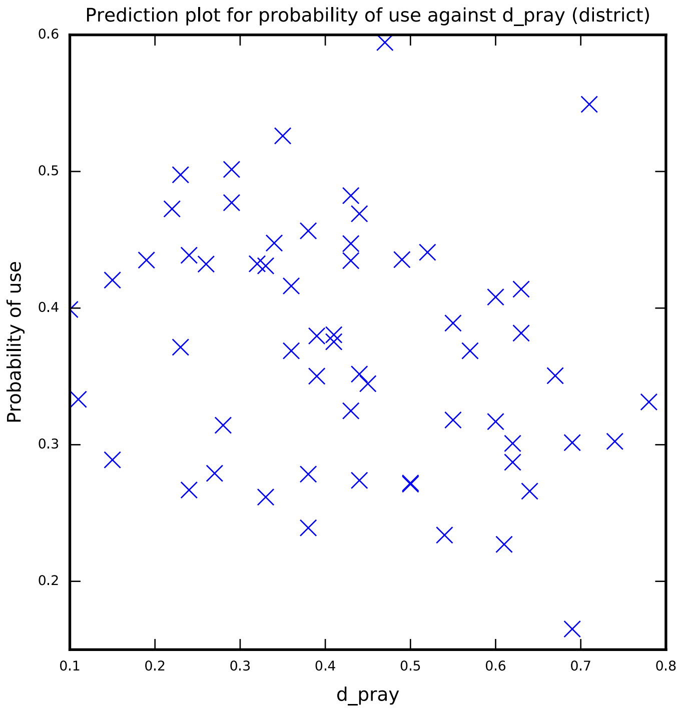
Our next step in modelling now that we have a set of potential predictors is to consider models for each predictor in turn along with a random intercept at each chosen classification from the best model in the last section. In the fixed part these models simply contain an intercept and the particular 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.

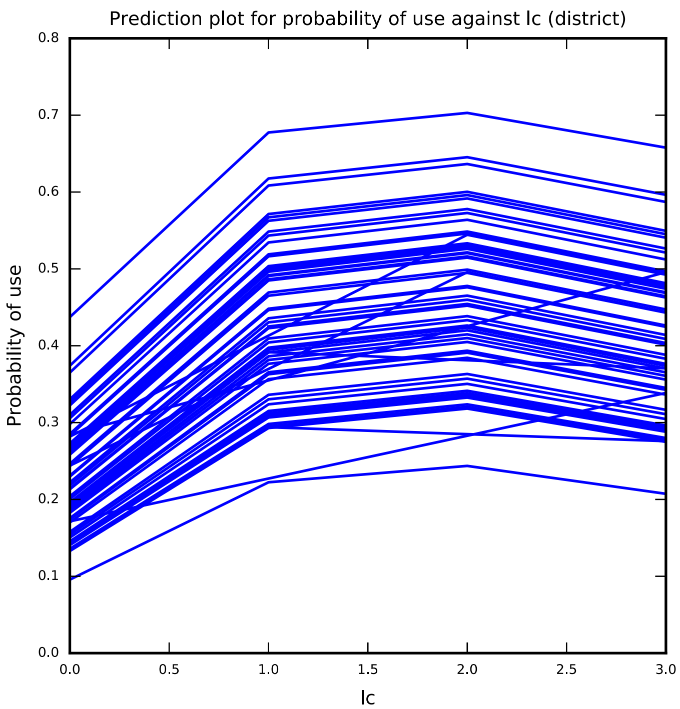
| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **age** | 0.00851 | 0.00532 | 1297 | 0.11 | N/S |
| **d_illit** | -2.803 | 0.566 | 284 | < 0.001 | *** |
| **d_pray** | -0.624 | 0.518 | 224 | 0.228 | N/S |
| **lc_1** | 0.994 | 0.148 | 776 | < 0.001 | *** |
| **lc_2** | 1.114 | 0.16 | 633 | | |
| **lc_3** | 0.906 | 0.133 | 657 | | |
| **urban_1** | 0.647 | 0.116 | 710 | < 0.001 | *** |
| **educ_2** | 0.42 | 0.153 | 892 | < 0.001 | *** |
| **educ_3** | 0.803 | 0.164 | 809 | | |
| **educ_4** | 1.321 | 0.141 | 811 | | |
| **hindu_1** | 0.241 | 0.151 | 881 | 0.112 | N/S |

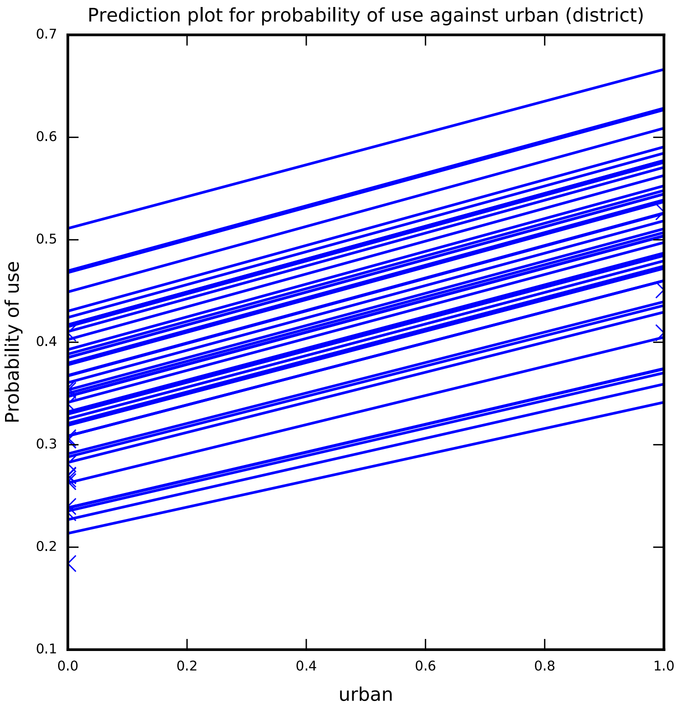
Which predictors we consider for the next stage of analysis will depend on their significance in the above table (but may in practice also depend on the size the effect and substantive interest of the variable though this is hard to automate). We will use a threshold on the p values associated with the predictors to decide whether to include the predictors in the next stage. Here we are currently using a threshold of 0.05. so the predictors to carry forward are: urban, lc, d_illit, and educ.

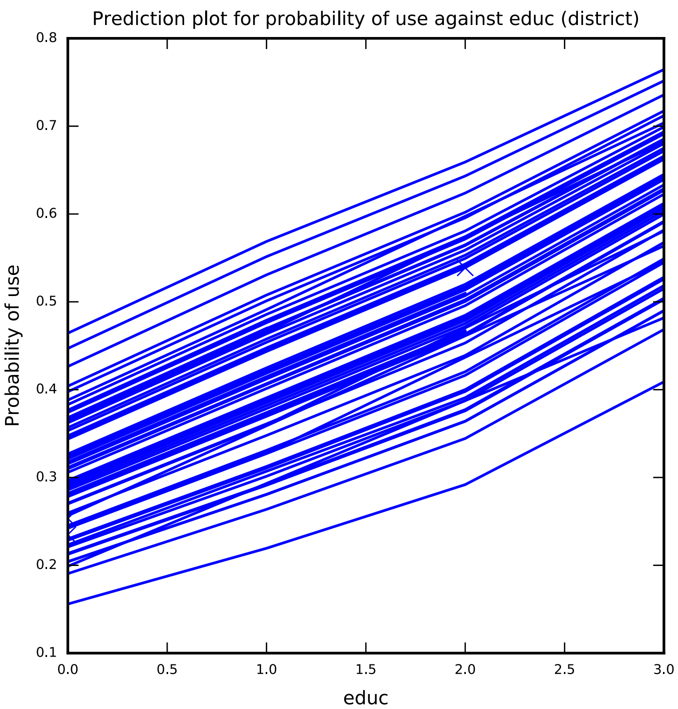

Prediction plot for probability of use against age (district)
Probability of use
age
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−15
−10
−5
0
5
10
15
20

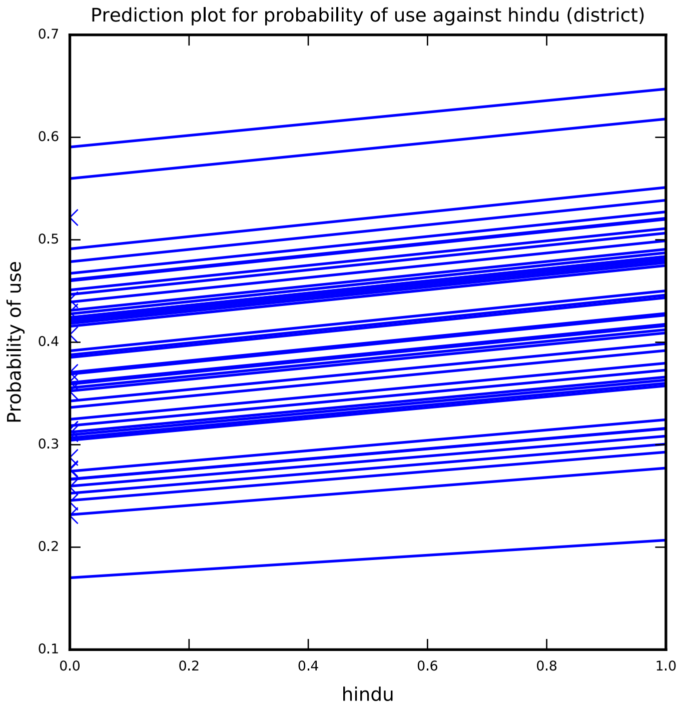

Prediction plot for probability of use against d_illit (district)
Probability of use
d_illit
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0


Prediction plot for probability of use against d_pray (district)
Probability of use
d_pray
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Prediction plot for probability of use against lc (district)
Probability of use
lc
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Prediction plot for probability of use against urban (district)
Probability of use
urban
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.2
0.4
0.6
0.8
1.0


Prediction plot for probability of use against educ (district)
Probability of use
educ
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Prediction plot for probability of use against hindu (district)
Probability of use
hindu
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.4
0.6
0.8
1.0

## 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:

| Variables | Correlation |
|---|---|
| **(d_illit, age)** | -0.037 |
| **(d_pray, age)** | 0.031 |
| **(d_pray, d_illit)** | -0.374 |
| **(lc_1, age)** | -0.206 |
| **(lc_1, d_illit)** | 0.018 |
| **(lc_1, d_pray)** | -0.046 |
| **(lc_2, age)** | 0.013 |
| **(lc_2, d_illit)** | -0.014 |
| **(lc_2, d_pray)** | 0.029 |
| **(lc_2, lc_1)** | -0.206 |
| **(lc_3, age)** | 0.632 |
| **(lc_3, d_illit)** | -0.016 |
| **(lc_3, d_pray)** | 0.053 |
| **(lc_3, lc_1)** | -0.374 |
| **(lc_3, lc_2)** | -0.343 |
| **(urban_1, age)** | -0.017 |
| **(urban_1, d_illit)** | -0.243 |
| **(urban_1, d_pray)** | -0.038 |
| **(urban_1, lc_1)** | 0.033 |
| **(urban_1, lc_2)** | -0.022 |
| **(urban_1, lc_3)** | -0.047 |
| **(educ_2, age)** | -0.024 |

| Variables | Correlation |
|---|---|
| **(educ_2, d_illit)** | -0.087 |
| **(educ_2, d_pray)** | -0.022 |
| **(educ_2, lc_1)** | -0.01 |
| **(educ_2, lc_2)** | 0.001 |
| **(educ_2, lc_3)** | 0.021 |
| **(educ_2, urban_1)** | -0.011 |
| **(educ_3, age)** | -0.049 |
| **(educ_3, d_illit)** | -0.103 |
| **(educ_3, d_pray)** | 0.07 |
| **(educ_3, lc_1)** | 0.037 |
| **(educ_3, lc_2)** | 0.013 |
| **(educ_3, lc_3)** | -0.022 |
| **(educ_3, urban_1)** | 0.016 |
| **(educ_3, educ_2)** | -0.122 |
| **(educ_4, age)** | -0.115 |
| **(educ_4, d_illit)** | -0.165 |
| **(educ_4, d_pray)** | 0.063 |
| **(educ_4, lc_1)** | 0.062 |
| **(educ_4, lc_2)** | 0.015 |
| **(educ_4, lc_3)** | -0.16 |
| **(educ_4, urban_1)** | 0.283 |
| **(educ_4, educ_2)** | -0.16 |

| Variables | Correlation |
| ---: | :---: |
| **(educ_4, educ_3)** | -0.139 |
| **(hindu_1, age)** | 0.011 |
| **(hindu_1, d_illit)** | 0.036 |
| **(hindu_1, d_pray)** | -0.044 |
| **(hindu_1, lc_1)** | 0.045 |
| **(hindu_1, lc_2)** | 0.022 |
| **(hindu_1, lc_3)** | -0.044 |
| **(hindu_1, urban_1)** | -0.001 |
| **(hindu_1, educ_2)** | -0.02 |
| **(hindu_1, educ_3)** | 0.036 |
| **(hindu_1, educ_4)** | 0.008 |

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 Wald tests to compare models. These work by looked at estimates and standard error matrices for each predictor to assess significance and run quicker than the alternative methods as they do not need to run submodels. It should be noted that the Wald test is an unusual choice for MCMC estimation even though we offer it here.

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

$$\mathrm{use}_i \sim Binomial(\mathrm{cons}_i, p_i), logit(p_i) = \beta_0 \mathrm{educ\_2}_i + \beta_1 \mathrm{educ\_3}_i + \beta_2 \mathrm{educ\_4}_i + \beta_3 \mathrm{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.42 | 0.153 | 892 | < 0.001 | *** |
| **educ_3** | 0.803 | 0.164 | 809 | | |
| **educ_4** | 1.321 | 0.141 | 811 | | |
| **Intercept** | -0.863 | 0.0857 | 257 | | |
| **Between district Variance** | 0.221 | 0.0777 | 313 | | |

Adding variable educ was a significant improvement and so we retain it in the model.

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

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i$$
$$+\beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.458 | 0.161 | 768 | < 0.001 | *** |
| **educ_3** | 0.861 | 0.172 | 676 | | |
| **educ_4** | 1.578 | 0.153 | 721 | | |
| **lc_1** | 1.095 | 0.157 | 769 | < 0.001 | *** |
| **lc_2** | 1.284 | 0.169 | 741 | | |
| **lc_3** | 1.206 | 0.142 | 673 | | |
| **Intercept** | -1.826 | 0.145 | 429 | | |
| **Between district Variance** | 0.257 | 0.091 | 257 | | |

Adding variable lc was a significant improvement and so we retain it in the model.

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

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i$$
$$+\beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{intercept}_i$$
$$+u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.434 | 0.16 | 704 | < 0.001 | *** |
| **educ_3** | 0.808 | 0.173 | 618 | | |
| **educ_4** | 1.438 | 0.154 | 635 | | |
| **lc_1** | 1.094 | 0.168 | 702 | < 0.001 | *** |
| **lc_2** | 1.303 | 0.163 | 728 | | |
| **lc_3** | 1.203 | 0.137 | 629 | | |
| **urban_1** | 0.422 | 0.124 | 486 | < 0.001 | *** |
| **Intercept** | -1.896 | 0.137 | 467 | | |
| **Between district Variance** | 0.221 | 0.0796 | 203 | | |

Adding variable urban was a significant improvement and so we retain it in the model.

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

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.399 | 0.155 | 727 | < 0.001 | *** |
| **educ_3** | 0.768 | 0.175 | 759 | | |
| **educ_4** | 1.424 | 0.163 | 570 | | |
| **lc_1** | 1.102 | 0.163 | 694 | < 0.001 | *** |
| **lc_2** | 1.299 | 0.168 | 729 | | |
| **lc_3** | 1.194 | 0.142 | 682 | | |
| **urban_1** | 0.39 | 0.122 | 690 | 0.001 | ** |
| **d_illit** | -1.774 | 0.618 | 283 | 0.004 | ** |
| **Intercept** | -0.766 | 0.429 | 248 | | |
| **Between district Variance** | 0.208 | 0.0792 | 223 | | |

Adding variable d_illit was a significant improvement and so we retain it in the model.

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

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0 \text{educ\_2}_i + \beta_1 \text{educ\_3}_i + \beta_2 \text{educ\_4}_i + \beta_3 \text{lc\_1}_i + \beta_4 \text{lc\_2}_i + \beta_5 \text{lc\_3}_i + \beta_6 \text{urban\_1}_i + \beta_7 \text{d\_illit}_i + \beta_8 \text{age}_i + \beta_9 \text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.357 | 0.156 | 1238 | < 0.001 | *** |
| **educ_3** | 0.727 | 0.174 | 1198 | | |
| **educ_4** | 1.397 | 0.159 | 1124 | | |
| **lc_1** | 1.198 | 0.165 | 1195 | < 0.001 | *** |
| **lc_2** | 1.491 | 0.186 | 1139 | | |
| **lc_3** | 1.562 | 0.189 | 1117 | | |
| **urban_1** | 0.403 | 0.133 | 842 | 0.002 | ** |
| **d_illit** | -1.881 | 0.649 | 257 | 0.004 | ** |
| **age** | -0.0242 | 0.00851 | 1197 | 0.004 | ** |
| **Intercept** | -0.865 | 0.437 | 305 | | |
| **Between district Variance** | 0.197 | 0.0755 | 297 | | |

Adding variable age was a significant improvement and so we retain it in the model.

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

$$\mathrm{use}_i \sim Binomial(\mathrm{cons}_i, p_i), logit(p_i) = \beta_0 \mathrm{educ\_2}_i + \beta_1 \mathrm{educ\_3}_i + \beta_2 \mathrm{educ\_4}_i + \beta_3 \mathrm{lc\_1}_i + \beta_4 \mathrm{lc\_2}_i + \beta_5 \mathrm{lc\_3}_i + \beta_6 \mathrm{urban\_1}_i + \beta_7 \mathrm{d\_illit}_i + \beta_8 \mathrm{age}_i + \beta_9 \mathrm{hindu\_1}_i + \beta_{10} \mathrm{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.362 | 0.15 | 851 | < 0.001 | *** |
| **educ_3** | 0.72 | 0.172 | 1013 | | |
| **educ_4** | 1.39 | 0.154 | 902 | | |
| **lc_1** | 1.206 | 0.161 | 925 | < 0.001 | *** |
| **lc_2** | 1.497 | 0.183 | 904 | | |
| **lc_3** | 1.581 | 0.183 | 897 | | |
| **urban_1** | 0.409 | 0.127 | 713 | 0.001 | ** |
| **d_illit** | -1.844 | 0.579 | 387 | 0.001 | ** |
| **age** | -0.0254 | 0.00802 | 996 | 0.002 | ** |
| **hindu_1** | 0.253 | 0.16 | 782 | 0.114 | N/S |
| **Intercept** | -0.947 | 0.392 | 433 | | |
| **Between district Variance** | 0.194 | 0.0697 | 232 | | |

Adding variable hindu 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 d_pray to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.343 | 0.157 | 2134 | < 0.001 | *** |
| **educ_3** | 0.736 | 0.169 | 2142 | | |
| **educ_4** | 1.421 | 0.161 | 1911 | | |
| **lc_1** | 1.206 | 0.165 | 2086 | < 0.001 | *** |
| **lc_2** | 1.514 | 0.183 | 2024 | | |
| **lc_3** | 1.589 | 0.189 | 2033 | | |
| **urban_1** | 0.35 | 0.124 | 1757 | 0.005 | ** |
| **d_illit** | -2.782 | 0.592 | 1038 | < 0.001 | *** |
| **age** | -0.0254 | 0.00828 | 2186 | 0.002 | ** |
| **d_pray** | -2.071 | 0.504 | 820 | < 0.001 | *** |
| **Intercept** | 0.587 | 0.521 | 969 | | |
| **Between district Variance** | 0.0992 | 0.0555 | 234 | | |

Adding variable d_pray was a significant improvement and so we retain it 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.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.335 | 0.158 | 3488 | < 0.001 | *** |
| **educ_3** | 0.724 | 0.178 | 3591 | | |
| **educ_4** | 1.342 | 0.16 | 3473 | | |
| **lc_1** | 0.933 | 0.17 | 3609 | < 0.001 | *** |
| **lc_2** | 1.088 | 0.198 | 3448 | | |
| **lc_3** | 1.172 | 0.201 | 3468 | | |
| **urban_1** | 0.331 | 0.125 | 3110 | 0.008 | ** |
| **d_illit** | -2.859 | 0.617 | 1392 | < 0.001 | *** |
| **age** | 0.00247 | 0.0097 | 3472 | 0.799 | N/S |
| **d_pray** | -2.101 | 0.502 | 1422 | < 0.001 | *** |
| **age_X_age** | -0.00411 | 0.000746 | 3609 | < 0.001 | *** |
| **Intercept** | 1.264 | 0.554 | 1380 | | |
| **Between district Variance** | 0.104 | 0.0598 | 225 | | |

Adding variable age_X_age significantly improved the model and so is retained in the model.

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

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.35 | 0.157 | 4939 | < 0.001 | *** |
| **educ_3** | 0.71 | 0.175 | 4984 | | |
| **educ_4** | 1.334 | 0.16 | 4864 | | |
| **lc_1** | 0.929 | 0.171 | 4704 | < 0.001 | *** |
| **lc_2** | 1.065 | 0.192 | 4821 | | |
| **lc_3** | 1.159 | 0.198 | 4879 | | |
| **urban_1** | 0.328 | 0.123 | 4227 | 0.008 | ** |
| **d_illit** | -2.917 | 0.599 | 2229 | < 0.001 | *** |
| **age** | 0.0383 | 0.0199 | 4910 | 0.054 | N/S |
| **d_pray** | -2.091 | 0.509 | 1734 | < 0.001 | *** |
| **age_X_age** | -0.00412 | 0.00075 | 4646 | < 0.001 | *** |
| **d_pray_X_age** | -0.0855 | 0.0417 | 4755 | 0.04 | * |
| **Intercept** | 1.31 | 0.54 | 2244 | | |
| **Between district Variance** | 0.1 | 0.0609 | 287 | | |

Adding variable d_pray_X_age significantly improved the model and so is retained in the model.

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

$$
\begin{aligned}
&\mathrm{use}_i \sim Binomial(\mathrm{cons}_i, p_i), logit(p_i) = \beta_0 \mathrm{educ\_2}_i + \beta_1 \mathrm{educ\_3}_i \\
&+ \beta_2 \mathrm{educ\_4}_i + \beta_3 \mathrm{lc\_1}_i + \beta_4 \mathrm{lc\_2}_i + \beta_5 \mathrm{lc\_3}_i + \beta_6 \mathrm{urban\_1}_i + \beta_7 \mathrm{d\_illit}_i + \beta_8 \mathrm{age}_i \\
&+ \beta_9 \mathrm{d\_pray}_i + \beta_{10} \mathrm{age\_X\_age}_i + \beta_{11} \mathrm{d\_pray\_X\_age}_i \\
&+ \beta_{12} \mathrm{urban\_1\_X\_educ\_2}_i + \beta_{13} \mathrm{urban\_1\_X\_educ\_3}_i \\
&+ \beta_{14} \mathrm{urban\_1\_X\_educ\_4}_i + \beta_{15} \mathrm{intercept}_i + u^{(2)}_{0,district_i}
\end{aligned}
$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.482 | 0.186 | 1701 | < 0.001 | *** |
| **educ_3** | 0.757 | 0.21 | 1651 | | |
| **educ_4** | 1.024 | 0.221 | 1797 | | |
| **lc_1** | 0.951 | 0.17 | 1729 | < 0.001 | *** |
| **lc_2** | 1.103 | 0.194 | 1669 | | |
| **lc_3** | 1.191 | 0.197 | 1738 | | |
| **urban_1** | 0.287 | 0.164 | 1427 | 0.08 | N/S |
| **d_illit** | -2.875 | 0.613 | 658 | < 0.001 | *** |
| **age** | 0.0385 | 0.0207 | 1717 | 0.063 | N/S |
| **d_pray** | -2.027 | 0.498 | 626 | < 0.001 | *** |
| **age_X_age** | -0.00411 | 0.000757 | 1618 | < 0.001 | *** |
| **d_pray_X_age** | -0.0879 | 0.0429 | 1688 | 0.04 | * |
| **urban_1_X_educ_2** | -0.431 | 0.351 | 1652 | 0.092 | N/S |
| **urban_1_X_educ_3** | -0.11 | 0.383 | 1656 | | |
| **urban_1_X_educ_4** | 0.59 | 0.317 | 1651 | | |
| **Intercept** | 1.232 | 0.535 | 721 | | |
| **Between district Variance** | 0.119 | 0.0603 | 215 | | |

Adding variable urban_X_educ 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 d_illit_X_urban to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_illit\_X\_urban\_1}_i + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.359 | 0.155 | 2493 | < 0.001 | *** |
| **educ_3** | 0.718 | 0.176 | 2336 | | |
| **educ_4** | 1.334 | 0.159 | 2316 | | |
| **lc_1** | 0.946 | 0.171 | 2336 | < 0.001 | *** |
| **lc_2** | 1.083 | 0.196 | 2417 | | |
| **lc_3** | 1.175 | 0.201 | 2328 | | |
| **urban_1** | 1.203 | 0.616 | 1623 | 0.051 | N/S |
| **d_illit** | -2.492 | 0.674 | 1121 | < 0.001 | *** |
| **age** | 0.0386 | 0.0198 | 2476 | 0.052 | N/S |
| **d_pray** | -2.104 | 0.502 | 823 | < 0.001 | *** |
| **age_X_age** | -0.00407 | 0.000757 | 2435 | < 0.001 | *** |
| **d_pray_X_age** | -0.0866 | 0.0413 | 2465 | 0.036 | * |
| **d_illit_X_urban_1** | -1.448 | 0.996 | 1671 | 0.146 | N/S |
| **Intercept** | 1.019 | 0.574 | 1038 | | |
| **Between district Variance** | 0.11 | 0.0641 | 205 | | |

Adding variable d_illit_X_urban 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 age_X_d_illit to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i$$
$$+\beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i$$
$$+\beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{age\_X\_d\_illit}_i$$
$$+\beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.353 | 0.156 | 4422 | < 0.001 | *** |
| **educ_3** | 0.716 | 0.177 | 4330 | | |
| **educ_4** | 1.34 | 0.16 | 4161 | | |
| **lc_1** | 0.93 | 0.169 | 4374 | < 0.001 | *** |
| **lc_2** | 1.073 | 0.193 | 4482 | | |
| **lc_3** | 1.157 | 0.197 | 4461 | | |
| **urban_1** | 0.323 | 0.127 | 3809 | 0.011 | * |
| **d_illit** | -2.942 | 0.604 | 1872 | < 0.001 | *** |
| **age** | 0.0137 | 0.0478 | 3707 | 0.774 | N/S |
| **d_pray** | -2.078 | 0.513 | 1503 | < 0.001 | *** |
| **age_X_age** | -0.00414 | 0.000746 | 3864 | < 0.001 | *** |
| **d_pray_X_age** | -0.0729 | 0.0471 | 3969 | 0.122 | N/S |
| **age_X_d_illit** | 0.0323 | 0.0558 | 3668 | 0.562 | N/S |
| **Intercept** | 1.322 | 0.537 | 1798 | | |
| **Between district Variance** | 0.107 | 0.0639 | 204 | | |

Adding variable age_X_d_illit 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 age_X_lc to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{age\_X\_lc\_1}_i + \beta_{13}\text{age\_X\_lc\_2}_i + \beta_{14}\text{age\_X\_lc\_3}_i + \beta_{15}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.344 | 0.163 | 1915 | < 0.001 | *** |
| **educ_3** | 0.708 | 0.178 | 1935 | | |
| **educ_4** | 1.325 | 0.164 | 1686 | | |
| **lc_1** | 1.268 | 0.265 | 1377 | < 0.001 | *** |
| **lc_2** | 1.399 | 0.257 | 1371 | | |
| **lc_3** | 1.332 | 0.25 | 1399 | | |
| **urban_1** | 0.338 | 0.13 | 1576 | 0.009 | ** |
| **d_illit** | -2.958 | 0.615 | 958 | < 0.001 | *** |
| **age** | -0.0146 | 0.0299 | 1298 | 0.625 | N/S |
| **d_pray** | -2.068 | 0.522 | 759 | < 0.001 | *** |
| **age_X_age** | -0.00635 | 0.00114 | 1701 | < 0.001 | *** |
| **d_pray_X_age** | -0.0935 | 0.0423 | 1894 | 0.027 | * |
| **age_X_lc_1** | 0.0531 | 0.0326 | 1427 | 0.06 | N/S |
| **age_X_lc_2** | 0.0763 | 0.0342 | 1450 | | |
| **age_X_lc_3** | 0.0972 | 0.036 | 1384 | | |
| **Intercept** | 1.077 | 0.555 | 1014 | | |
| **Between district Variance** | 0.119 | 0.0574 | 222 | | |

Adding variable age_X_lc 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 d_pray_X_educ to the current model.

$$\mathrm{use}_i \sim Binomial(\mathrm{cons}_i, p_i), logit(p_i) = \beta_0 \mathrm{educ\_2}_i + \beta_1 \mathrm{educ\_3}_i$$
$$+ \beta_2 \mathrm{educ\_4}_i + \beta_3 \mathrm{lc\_1}_i + \beta_4 \mathrm{lc\_2}_i + \beta_5 \mathrm{lc\_3}_i + \beta_6 \mathrm{urban\_1}_i + \beta_7 \mathrm{d\_illit}_i + \beta_8 \mathrm{age}_i$$
$$+ \beta_9 \mathrm{d\_pray}_i + \beta_{10} \mathrm{age\_X\_age}_i + \beta_{11} \mathrm{d\_pray\_X\_age}_i$$
$$+ \beta_{12} \mathrm{d\_pray\_X\_educ\_2}_i + \beta_{13} \mathrm{d\_pray\_X\_educ\_3}_i$$
$$+ \beta_{14} \mathrm{d\_pray\_X\_educ\_4}_i + \beta_{15} \mathrm{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.0612 | 0.479 | 4355 | 0.311 | N/S |
| **educ_3** | -0.27 | 0.539 | 4368 | | |
| **educ_4** | 0.825 | 0.481 | 4135 | | |
| **lc_1** | 0.94 | 0.173 | 4382 | < 0.001 | *** |
| **lc_2** | 1.063 | 0.196 | 4387 | | |
| **lc_3** | 1.157 | 0.202 | 4360 | | |
| **urban_1** | 0.342 | 0.125 | 4017 | 0.006 | ** |
| **d_illit** | -2.877 | 0.614 | 2029 | < 0.001 | *** |
| **age** | 0.0362 | 0.0202 | 4360 | 0.073 | N/S |
| **d_pray** | -2.541 | 0.587 | 1804 | < 0.001 | *** |
| **age_X_age** | -0.00416 | 0.000752 | 4164 | < 0.001 | *** |
| **d_pray_X_age** | -0.079 | 0.0426 | 4149 | 0.064 | N/S |
| **d_pray_X_educ_2** | 0.704 | 1.087 | 4302 | 0.216 | N/S |
| **d_pray_X_educ_3** | 2.222 | 1.139 | 4264 | | |
| **d_pray_X_educ_4** | 1.165 | 1.011 | 4109 | | |
| **Intercept** | 1.46 | 0.561 | 2065 | | |
| **Between district Variance** | 0.11 | 0.0649 | 211 | | |

Adding variable d_pray_X_educ 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 age_X_educ to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{age\_X\_educ\_2}_i + \beta_{13}\text{age\_X\_educ\_3}_i + \beta_{14}\text{age\_X\_educ\_4}_i + \beta_{15}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.358 | 0.16 | 1731 | < 0.001 | *** |
| **educ_3** | 0.691 | 0.179 | 1910 | | |
| **educ_4** | 1.35 | 0.168 | 1877 | | |
| **lc_1** | 0.937 | 0.169 | 2050 | < 0.001 | *** |
| **lc_2** | 1.079 | 0.193 | 2020 | | |
| **lc_3** | 1.168 | 0.203 | 1987 | | |
| **urban_1** | 0.342 | 0.124 | 1695 | 0.006 | ** |
| **d_illit** | -2.845 | 0.588 | 992 | < 0.001 | *** |
| **age** | 0.0431 | 0.0205 | 1774 | 0.035 | * |
| **d_pray** | -2.065 | 0.504 | 795 | < 0.001 | *** |
| **age_X_age** | -0.00429 | 0.000754 | 1887 | < 0.001 | *** |
| **d_pray_X_age** | -0.0768 | 0.042 | 1747 | 0.067 | N/S |
| **age_X_educ_2** | -0.0212 | 0.0186 | 1922 | 0.194 | N/S |
| **age_X_educ_3** | -0.0418 | 0.0211 | 1811 | | |
| **age_X_educ_4** | -0.00697 | 0.0195 | 1970 | | |
| **Intercept** | 1.241 | 0.53 | 947 | | |
| **Between district Variance** | 0.107 | 0.0566 | 225 | | |

Adding variable age_X_educ 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 d_pray_X_lc to the current model.

$$\begin{aligned}&\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i \\ &+\beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i \\ &+\beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_pray\_X\_lc\_1}_i \\ &+\beta_{13}\text{d\_pray\_X\_lc\_2}_i + \beta_{14}\text{d\_pray\_X\_lc\_3}_i + \beta_{15}\text{intercept}_i + u^{(2)}_{0,district_i}\end{aligned}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.354 | 0.159 | 1948 | < 0.001 | *** |
| **educ_3** | 0.723 | 0.174 | 1969 | | |
| **educ_4** | 1.331 | 0.161 | 1731 | | |
| **lc_1** | 1.462 | 0.491 | 1879 | 0.009 | ** |
| **lc_2** | 1.255 | 0.533 | 1888 | | |
| **lc_3** | 1.628 | 0.554 | 1872 | | |
| **urban_1** | 0.328 | 0.129 | 1703 | 0.011 | * |
| **d_illit** | -2.964 | 0.62 | 850 | < 0.001 | *** |
| **age** | 0.028 | 0.0258 | 1892 | 0.278 | N/S |
| **d_pray** | -1.303 | 0.972 | 1166 | 0.18 | N/S |
| **age_X_age** | -0.00417 | 0.000771 | 1920 | < 0.001 | *** |
| **d_pray_X_age** | -0.0605 | 0.0564 | 1859 | 0.283 | N/S |
| **d_pray_X_lc_1** | -1.299 | 1.132 | 1798 | 0.643 | N/S |
| **d_pray_X_lc_2** | -0.454 | 1.16 | 1822 | | |
| **d_pray_X_lc_3** | -1.114 | 1.227 | 1692 | | |
| **Intercept** | 1.009 | 0.656 | 1027 | | |
| **Between district Variance** | 0.111 | 0.0625 | 212 | | |

Adding variable d_pray_X_lc 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 d_pray_X_d_pray to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_pray\_X\_d\_pray}_i + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.349 | 0.161 | 4834 | < 0.001 | *** |
| **educ_3** | 0.715 | 0.175 | 5028 | | |
| **educ_4** | 1.332 | 0.159 | 4567 | | |
| **lc_1** | 0.932 | 0.173 | 4871 | < 0.001 | *** |
| **lc_2** | 1.067 | 0.195 | 4700 | | |
| **lc_3** | 1.165 | 0.202 | 4696 | | |
| **urban_1** | 0.328 | 0.128 | 3948 | 0.01 | * |
| **d_illit** | -3.031 | 0.615 | 2219 | < 0.001 | *** |
| **age** | 0.0381 | 0.0204 | 4825 | 0.061 | N/S |
| **d_pray** | -3.722 | 2.503 | 2409 | 0.137 | N/S |
| **age_X_age** | -0.0041 | 0.000736 | 4762 | < 0.001 | *** |
| **d_pray_X_age** | -0.0854 | 0.0422 | 4749 | 0.043 | * |
| **d_pray_X_d_pray** | 1.809 | 2.714 | 2373 | 0.505 | N/S |
| **Intercept** | 1.698 | 0.79 | 2320 | | |
| **Between district Variance** | 0.102 | 0.06 | 204 | | |

Adding variable d_pray_X_d_pray 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 age_X_urban to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{age\_X\_urban\_1}_i + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.349 | 0.16 | 2371 | < 0.001 | *** |
| **educ_3** | 0.73 | 0.176 | 2442 | | |
| **educ_4** | 1.329 | 0.159 | 2367 | | |
| **lc_1** | 0.925 | 0.17 | 2406 | < 0.001 | *** |
| **lc_2** | 1.067 | 0.196 | 2367 | | |
| **lc_3** | 1.161 | 0.198 | 2182 | | |
| **urban_1** | 0.329 | 0.128 | 1928 | 0.01 | * |
| **d_illit** | -2.885 | 0.593 | 1175 | < 0.001 | *** |
| **age** | 0.0432 | 0.0205 | 2468 | 0.035 | * |
| **d_pray** | -2.048 | 0.507 | 859 | < 0.001 | *** |
| **age_X_age** | -0.00417 | 0.000757 | 2365 | < 0.001 | *** |
| **d_pray_X_age** | -0.0865 | 0.0422 | 2390 | 0.04 | * |
| **age_X_urban_1** | -0.0142 | 0.0132 | 2404 | 0.28 | N/S |
| **Intercept** | 1.272 | 0.526 | 1145 | | |
| **Between district Variance** | 0.107 | 0.0582 | 208 | | |

Adding variable age_X_urban 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 lc_X_educ to the current model.

$$
\begin{aligned}
&\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i \\
&+\beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i \\
&+\beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{lc\_1\_X\_educ\_2}_i \\
&+\beta_{13}\text{lc\_1\_X\_educ\_3}_i + \beta_{14}\text{lc\_1\_X\_educ\_4}_i + \beta_{15}\text{lc\_2\_X\_educ\_2}_i \\
&+\beta_{16}\text{lc\_2\_X\_educ\_3}_i + \beta_{17}\text{lc\_2\_X\_educ\_4}_i + \beta_{18}\text{lc\_3\_X\_educ\_2}_i \\
&+\beta_{19}\text{lc\_3\_X\_educ\_3}_i + \beta_{20}\text{lc\_3\_X\_educ\_4}_i + \beta_{21}\text{intercept}_i + u^{(2)}_{0,district_i}
\end{aligned}
$$

| Variable | Coefficient | SD | ESS | p value | Significance |
| --- | --- | --- | --- | --- | --- |
| **educ_2** | -0.161 | 0.401 | 1184 | < 0.001 | *** |
| **educ_3** | 1.028 | 0.37 | 1418 | | |
| **educ_4** | 1.252 | 0.264 | 1610 | | |
| **lc_1** | 0.869 | 0.231 | 1391 | < 0.001 | *** |
| **lc_2** | 0.904 | 0.25 | 1532 | | |
| **lc_3** | 1.171 | 0.23 | 1554 | | |
| **urban_1** | 0.344 | 0.129 | 1513 | 0.007 | ** |
| **d_illit** | -2.933 | 0.6 | 695 | < 0.001 | *** |
| **age** | 0.0362 | 0.0198 | 1676 | 0.068 | N/S |
| **d_pray** | -2.111 | 0.506 | 619 | < 0.001 | *** |
| **age_X_age** | -0.00412 | 0.000749 | 1777 | < 0.001 | *** |
| **d_pray_X_age** | -0.0813 | 0.0417 | 1609 | 0.051 | N/S |
| **lc_1_X_educ_2** | 0.955 | 0.549 | 1375 | 0.558 | N/S |
| **lc_1_X_educ_3** | -0.582 | 0.522 | 1458 | | |
| **lc_1_X_educ_4** | 0.112 | 0.421 | 1572 | | |
| **lc_2_X_educ_2** | 0.87 | 0.555 | 1251 | | |
| **lc_2_X_educ_3** | 0.0147 | 0.577 | 1446 | | |
| **lc_2_X_educ_4** | 0.454 | 0.472 | 1637 | | |
| **lc_3_X_educ_2** | 0.373 | 0.461 | 1509 | | |
| **lc_3_X_educ_3** | -0.481 | 0.458 | 1479 | | |

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **lc_3_X_educ_4** | -0.0227 | 0.411 | 1625 | | |
| **Intercept** | 1.352 | 0.547 | 809 | | |
| **Between district Variance** | 0.102 | 0.0542 | 203 | | |

Adding variable lc_X_educ 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 d_pray_X_urban to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_pray\_X\_urban\_1}_i + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.355 | 0.156 | 2722 | < 0.001 | *** |
| **educ_3** | 0.726 | 0.174 | 2619 | | |
| **educ_4** | 1.344 | 0.162 | 2594 | | |
| **lc_1** | 0.932 | 0.172 | 2693 | < 0.001 | *** |
| **lc_2** | 1.074 | 0.196 | 2594 | | |
| **lc_3** | 1.168 | 0.196 | 2685 | | |
| **urban_1** | 0.109 | 0.362 | 2484 | 0.762 | N/S |
| **d_illit** | -2.891 | 0.614 | 1247 | < 0.001 | *** |
| **age** | 0.0391 | 0.0195 | 2735 | 0.046 | * |
| **d_pray** | -2.219 | 0.55 | 1138 | < 0.001 | *** |
| **age_X_age** | -0.00415 | 0.000749 | 2647 | < 0.001 | *** |
| **d_pray_X_age** | -0.087 | 0.0414 | 2612 | 0.036 | * |
| **d_pray_X_urban_1** | 0.544 | 0.835 | 2266 | 0.515 | N/S |
| **Intercept** | 1.339 | 0.556 | 1202 | | |
| **Between district Variance** | 0.115 | 0.0598 | 239 | | |

Adding variable d_pray_X_urban 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 d_illit_X_educ to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_illit\_X\_educ\_2}_i$$

$$+\beta_{13}\text{d\_illit\_X\_educ\_3}_i + \beta_{14}\text{d\_illit\_X\_educ\_4}_i + \beta_{15}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.513 | 0.761 | 2391 | 0.034 | * |
| **educ_3** | 1.982 | 0.878 | 2349 | | |
| **educ_4** | 1.686 | 0.727 | 2435 | | |
| **lc_1** | 0.935 | 0.167 | 2391 | < 0.001 | *** |
| **lc_2** | 1.069 | 0.193 | 2501 | | |
| **lc_3** | 1.155 | 0.197 | 2438 | | |
| **urban_1** | 0.329 | 0.127 | 2001 | 0.009 | ** |
| **d_illit** | -2.542 | 0.703 | 1153 | < 0.001 | *** |
| **age** | 0.0387 | 0.0203 | 2423 | 0.057 | N/S |
| **d_pray** | -2.071 | 0.531 | 686 | < 0.001 | *** |
| **age_X_age** | -0.00415 | 0.000762 | 2285 | < 0.001 | *** |
| **d_pray_X_age** | -0.0846 | 0.0423 | 2374 | 0.046 | * |
| **d_illit_X_educ_2** | -0.242 | 1.253 | 2343 | 0.53 | N/S |
| **d_illit_X_educ_3** | -2.13 | 1.449 | 2372 | | |
| **d_illit_X_educ_4** | -0.573 | 1.207 | 2404 | | |
| **Intercept** | 1.061 | 0.593 | 982 | | |
| **Between district Variance** | 0.113 | 0.0603 | 226 | | |

Adding variable d_illit_X_educ 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 d_illit_X_d_illit to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i \\ + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i \\ + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_illit\_X\_d\_illit}_i \\ + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---:|:---:|:---:|:---:|:---:|:---:|
| **educ_2** | 0.354 | 0.158 | 1978 | < 0.001 | *** |
| **educ_3** | 0.721 | 0.174 | 1883 | | |
| **educ_4** | 1.342 | 0.154 | 1995 | | |
| **lc_1** | 0.926 | 0.171 | 1945 | < 0.001 | *** |
| **lc_2** | 1.067 | 0.196 | 1955 | | |
| **lc_3** | 1.156 | 0.201 | 1928 | | |
| **urban_1** | 0.326 | 0.125 | 1543 | 0.009 | ** |
| **d_illit** | -4.772 | 3.812 | 743 | 0.211 | N/S |
| **age** | 0.0384 | 0.0196 | 1935 | 0.051 | N/S |
| **d_pray** | -2.066 | 0.515 | 651 | < 0.001 | *** |
| **age_X_age** | -0.00413 | 0.000744 | 1875 | < 0.001 | *** |
| **d_pray_X_age** | -0.0849 | 0.0413 | 1854 | 0.04 | * |
| **d_illit_X_d_illit** | 1.528 | 3.096 | 737 | 0.622 | N/S |
| **Intercept** | 1.836 | 1.23 | 736 | | |
| **Between district Variance** | 0.116 | 0.0606 | 217 | | |

Adding variable d_illit_X_d_illit 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 d_illit_X_lc to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0 \text{educ\_2}_i + \beta_1 \text{educ\_3}_i$$
$$+\beta_2 \text{educ\_4}_i + \beta_3 \text{lc\_1}_i + \beta_4 \text{lc\_2}_i + \beta_5 \text{lc\_3}_i + \beta_6 \text{urban\_1}_i + \beta_7 \text{d\_illit}_i + \beta_8 \text{age}_i$$
$$+\beta_9 \text{d\_pray}_i + \beta_{10} \text{age\_X\_age}_i + \beta_{11} \text{d\_pray\_X\_age}_i + \beta_{12} \text{d\_illit\_X\_lc\_1}_i$$
$$+\beta_{13} \text{d\_illit\_X\_lc\_2}_i + \beta_{14} \text{d\_illit\_X\_lc\_3}_i + \beta_{15} \text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.353 | 0.166 | 1932 | < 0.001 | *** |
| **educ_3** | 0.723 | 0.176 | 1896 | | |
| **educ_4** | 1.34 | 0.157 | 1952 | | |
| **lc_1** | 1.099 | 0.834 | 1830 | 0.462 | N/S |
| **lc_2** | 0.996 | 0.856 | 1927 | | |
| **lc_3** | 0.996 | 0.723 | 1826 | | |
| **urban_1** | 0.329 | 0.128 | 1718 | 0.01 | * |
| **d_illit** | -3.044 | 0.994 | 1183 | 0.002 | ** |
| **age** | 0.0364 | 0.0214 | 1887 | 0.09 | N/S |
| **d_pray** | -2.089 | 0.536 | 766 | < 0.001 | *** |
| **age_X_age** | -0.00413 | 0.000758 | 1844 | < 0.001 | *** |
| **d_pray_X_age** | -0.0812 | 0.0457 | 1834 | 0.076 | N/S |
| **d_illit_X_lc_1** | -0.262 | 1.337 | 1668 | 0.976 | N/S |
| **d_illit_X_lc_2** | 0.127 | 1.356 | 1857 | | |
| **d_illit_X_lc_3** | 0.291 | 1.143 | 1647 | | |
| **Intercept** | 1.367 | 0.724 | 1195 | | |
| **Between district Variance** | 0.115 | 0.0591 | 235 | | |

Adding variable d_illit_X_lc 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 urban_X_lc to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{urban\_1\_X\_lc\_1}_i + \beta_{13}\text{urban\_1\_X\_lc\_2}_i + \beta_{14}\text{urban\_1\_X\_lc\_3}_i + \beta_{15}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.346 | 0.157 | 1933 | < 0.001 | *** |
| **educ_3** | 0.721 | 0.173 | 1924 | | |
| **educ_4** | 1.334 | 0.157 | 1897 | | |
| **lc_1** | 0.98 | 0.216 | 1846 | < 0.001 | *** |
| **lc_2** | 1.148 | 0.229 | 1783 | | |
| **lc_3** | 1.191 | 0.225 | 1786 | | |
| **urban_1** | 0.409 | 0.238 | 1626 | 0.086 | N/S |
| **d_illit** | -2.911 | 0.597 | 924 | < 0.001 | *** |
| **age** | 0.0377 | 0.02 | 1950 | 0.06 | N/S |
| **d_pray** | -2.068 | 0.514 | 733 | < 0.001 | *** |
| **age_X_age** | -0.00415 | 0.000767 | 1718 | < 0.001 | *** |
| **d_pray_X_age** | -0.0845 | 0.0423 | 1854 | 0.046 | * |
| **urban_1_X_lc_1** | -0.131 | 0.335 | 2008 | 0.926 | N/S |
| **urban_1_X_lc_2** | -0.224 | 0.367 | 1670 | | |
| **urban_1_X_lc_3** | -0.0366 | 0.29 | 1899 | | |
| **Intercept** | 1.255 | 0.539 | 921 | | |
| **Between district Variance** | 0.108 | 0.0595 | 214 | | |

Adding variable urban_X_lc 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 d_pray_X_d_illit to the current model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{d\_pray\_X\_d\_illit}_i + \beta_{13}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.357 | 0.157 | 1971 | < 0.001 | *** |
| **educ_3** | 0.721 | 0.176 | 1881 | | |
| **educ_4** | 1.331 | 0.157 | 1881 | | |
| **lc_1** | 0.921 | 0.172 | 1926 | < 0.001 | *** |
| **lc_2** | 1.066 | 0.194 | 2003 | | |
| **lc_3** | 1.155 | 0.198 | 1974 | | |
| **urban_1** | 0.33 | 0.129 | 1628 | 0.011 | * |
| **d_illit** | -2.408 | 1.748 | 837 | 0.168 | N/S |
| **age** | 0.0389 | 0.0202 | 1984 | 0.054 | N/S |
| **d_pray** | -1.37 | 2.351 | 811 | 0.56 | N/S |
| **age_X_age** | -0.00415 | 0.000739 | 1994 | < 0.001 | *** |
| **d_pray_X_age** | -0.0867 | 0.0416 | 1953 | 0.037 | * |
| **d_pray_X_d_illit** | -1.014 | 3.433 | 843 | 0.768 | N/S |
| **Intercept** | 0.948 | 1.241 | 789 | | |
| **Between district Variance** | 0.114 | 0.0588 | 234 | | |

Adding variable d_pray_X_d_illit did not significantly improve the model, so we remove it from the model.

We have considered all interaction variables so now run our final model.

$$\text{use}_i \sim Binomial(\text{cons}_i, p_i), logit(p_i) = \beta_0\text{educ\_2}_i + \beta_1\text{educ\_3}_i + \beta_2\text{educ\_4}_i + \beta_3\text{lc\_1}_i + \beta_4\text{lc\_2}_i + \beta_5\text{lc\_3}_i + \beta_6\text{urban\_1}_i + \beta_7\text{d\_illit}_i + \beta_8\text{age}_i + \beta_9\text{d\_pray}_i + \beta_{10}\text{age\_X\_age}_i + \beta_{11}\text{d\_pray\_X\_age}_i + \beta_{12}\text{intercept}_i + u^{(2)}_{0,district_i}$$

| Variable | Coefficient | SD | ESS | p value | Significance |
|---|---|---|---|---|---|
| **educ_2** | 0.35 | 0.157 | 4939 | < 0.001 | *** |
| **educ_3** | 0.71 | 0.175 | 4984 | | |
| **educ_4** | 1.334 | 0.16 | 4864 | | |
| **lc_1** | 0.929 | 0.171 | 4704 | < 0.001 | *** |
| **lc_2** | 1.065 | 0.192 | 4821 | | |
| **lc_3** | 1.159 | 0.198 | 4879 | | |
| **urban_1** | 0.328 | 0.123 | 4227 | 0.008 | ** |
| **d_illit** | -2.917 | 0.599 | 2229 | < 0.001 | *** |
| **age** | 0.0383 | 0.0199 | 4910 | 0.054 | N/S |
| **d_pray** | -2.091 | 0.509 | 1734 | < 0.001 | *** |
| **age_X_age** | -0.00412 | 0.00075 | 4646 | < 0.001 | *** |
| **d_pray_X_age** | -0.0855 | 0.0417 | 4755 | 0.04 | * |
| **Intercept** | 1.31 | 0.54 | 2244 | | |
| **Between district Variance** | 0.1 | 0.0609 | 287 | | |

This is our final model.

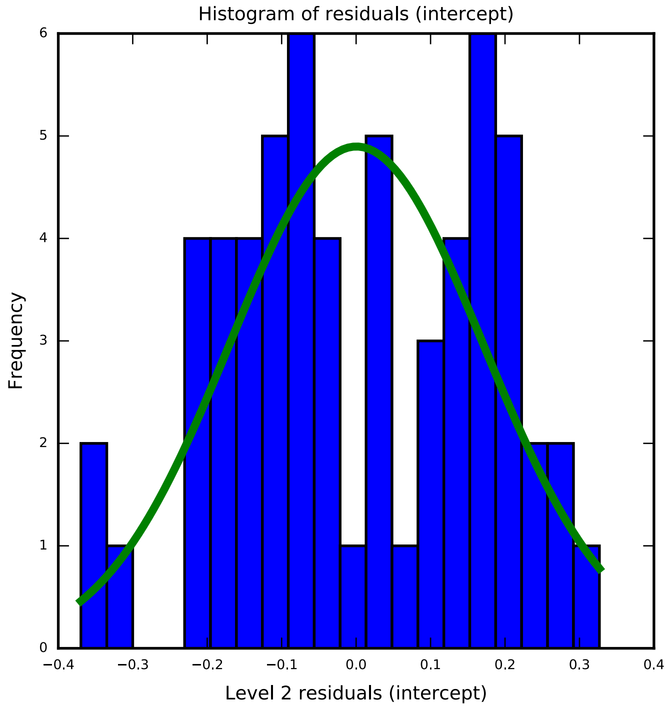
## Adding random slopes

You have chosen not to look at random slopes and so this page is blank.

# Analysing the residuals

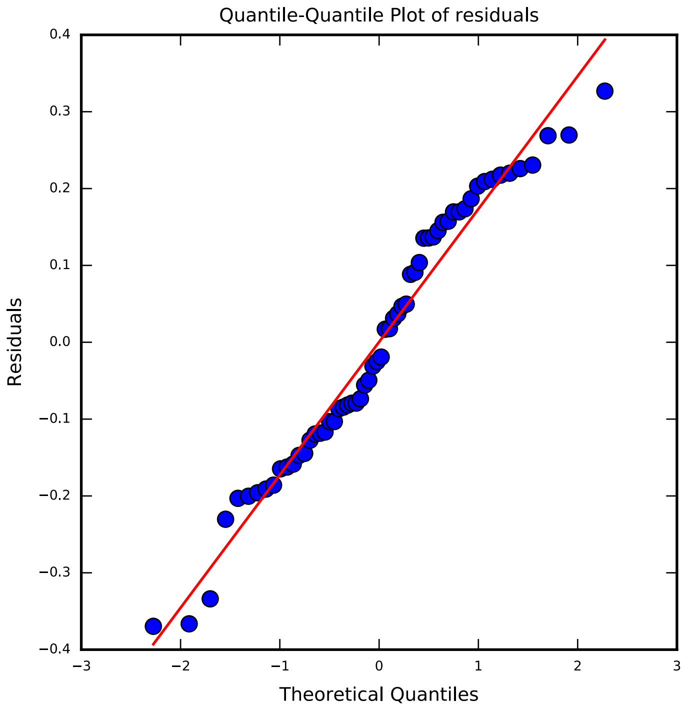
Here we look at the residuals from the model and plot them in various ways.

Next the level 2 residuals for intercept:




Here the distribution is reasonably symmetric with skewness value -0.108.

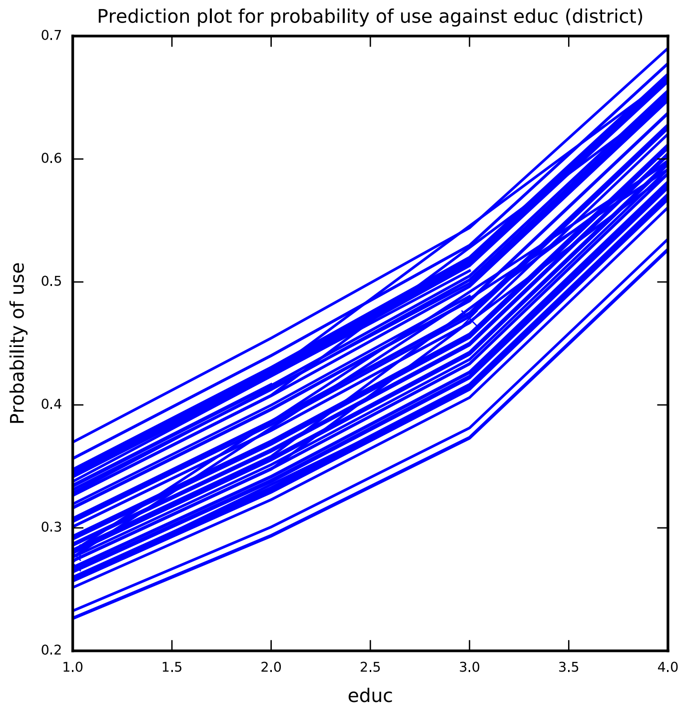
There are no obvious outliers in the residuals.

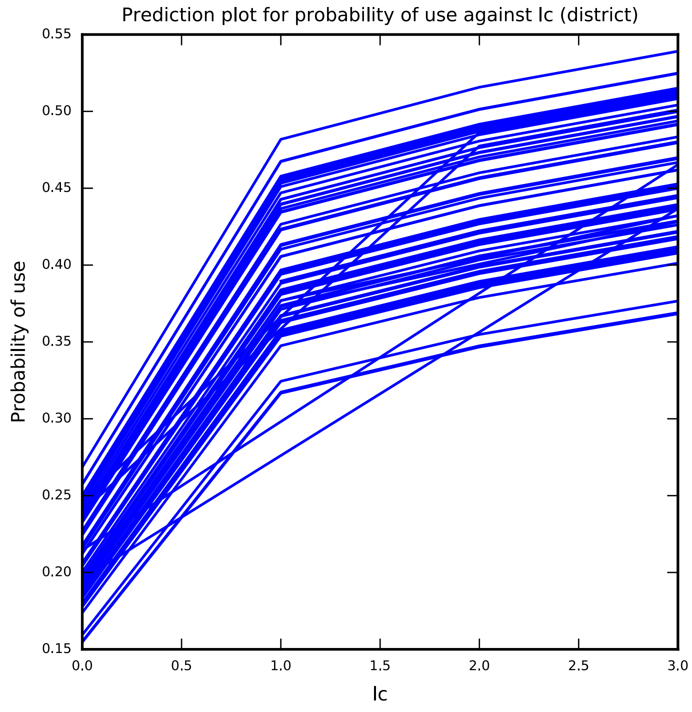


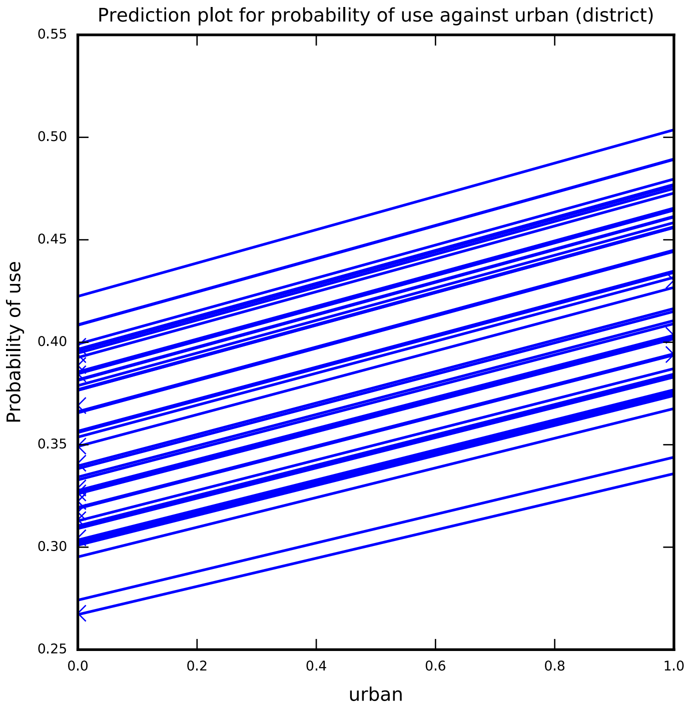

If the residuals are fairly normally distributed then the points in this graph should be close to the red line.

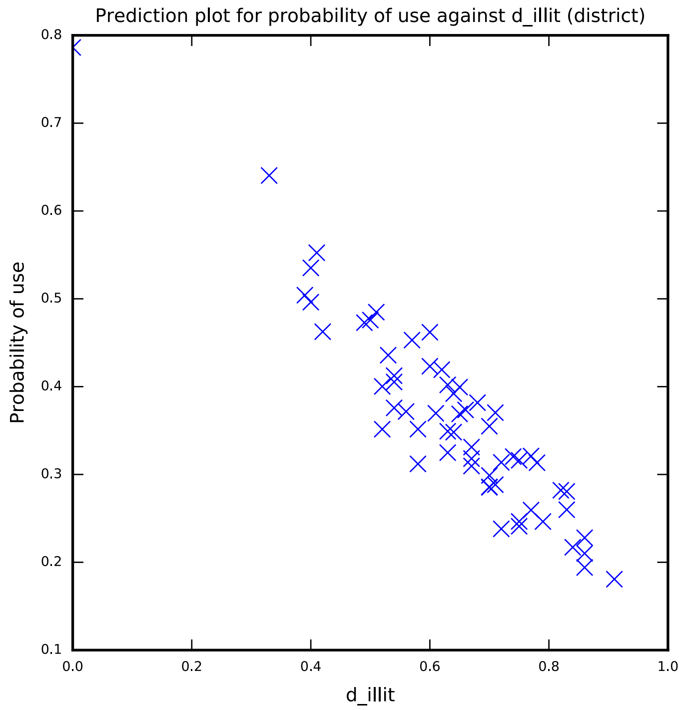
# Looking at predictions

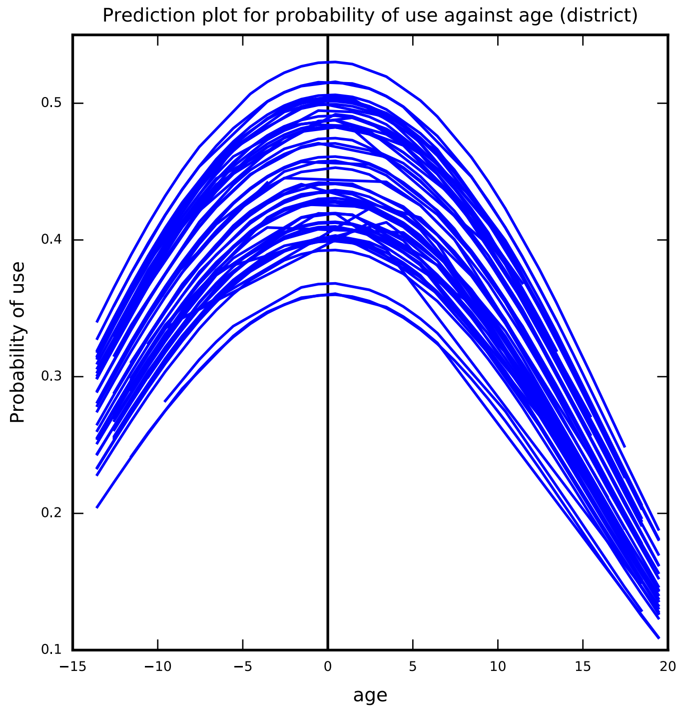
Having fitted a model with several predictors we might like to represent this model graphically. This is more difficult than when we have only one predictor and so for now we consider each predictor in turn and set all other predictors to their mean values.



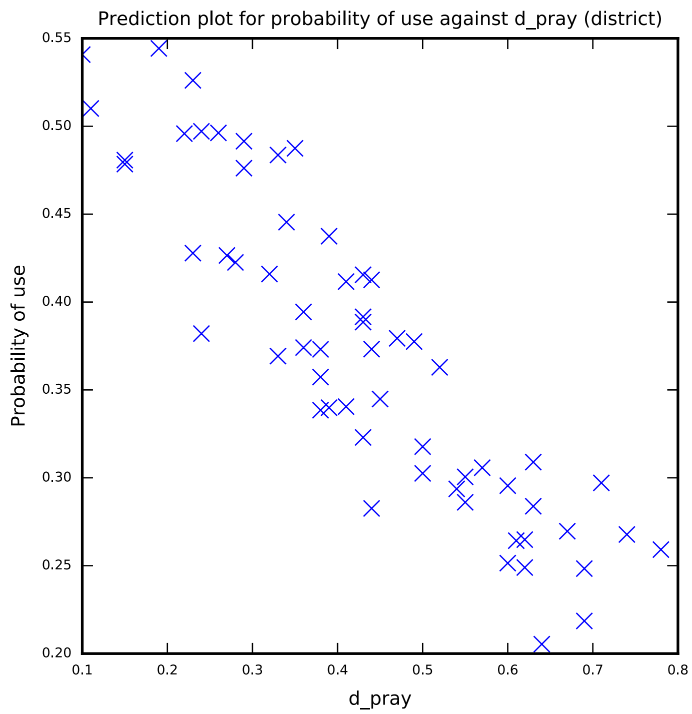




Prediction plot for probability of use against urban (district)
Probability of use
urban
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.0
0.2
0.4
0.6
0.8
1.0


Prediction plot for probability of use against d_illit (district)
Probability of use
d_illit
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0



Prediction plot for probability of use against age (district)
Probability of use
age
0.1
0.2
0.3
0.4
0.5
−15
−10
−5
0
5
10
15
20

Prediction plot for probability of use against d_pray (district)
Probability of use
d_pray
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8