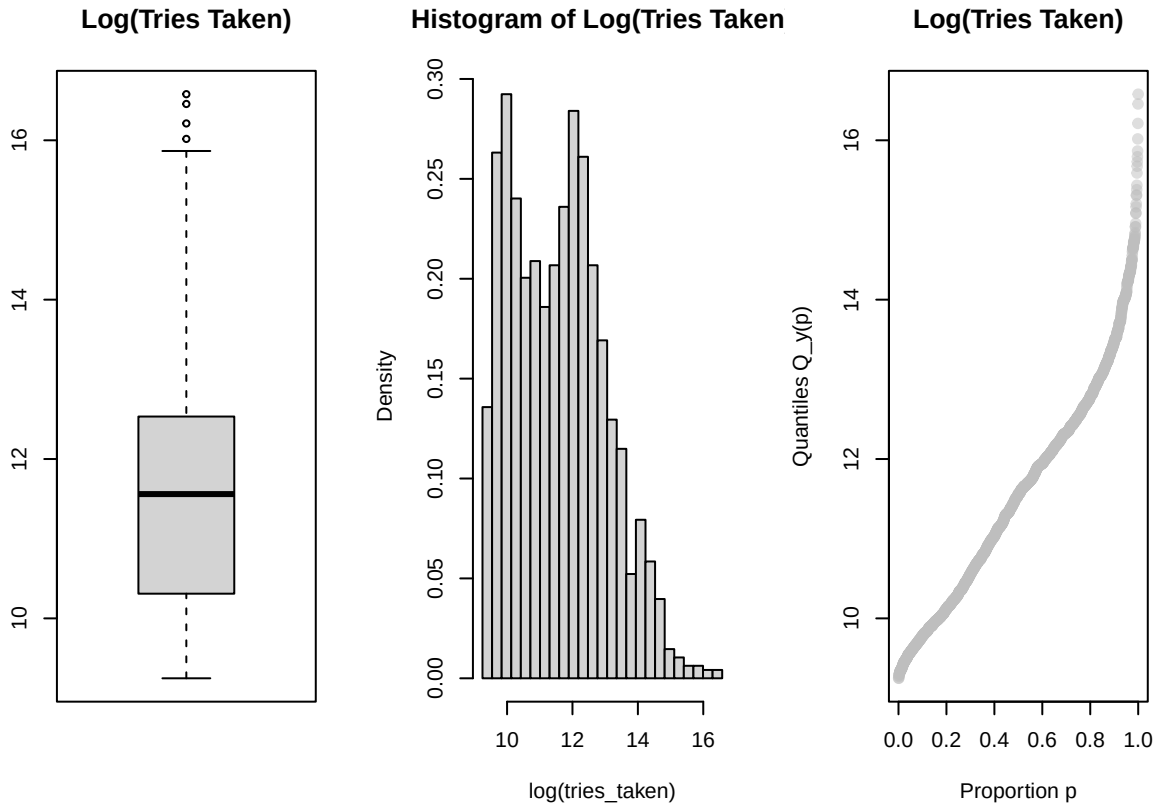
Graph SMM

We look at a subset of the Super Mario Maker Levels (SMML).

We consider the log(number of tries taken) for each level. We plot a boxplot, histogram and quantile plot for the log(number of tries taken).

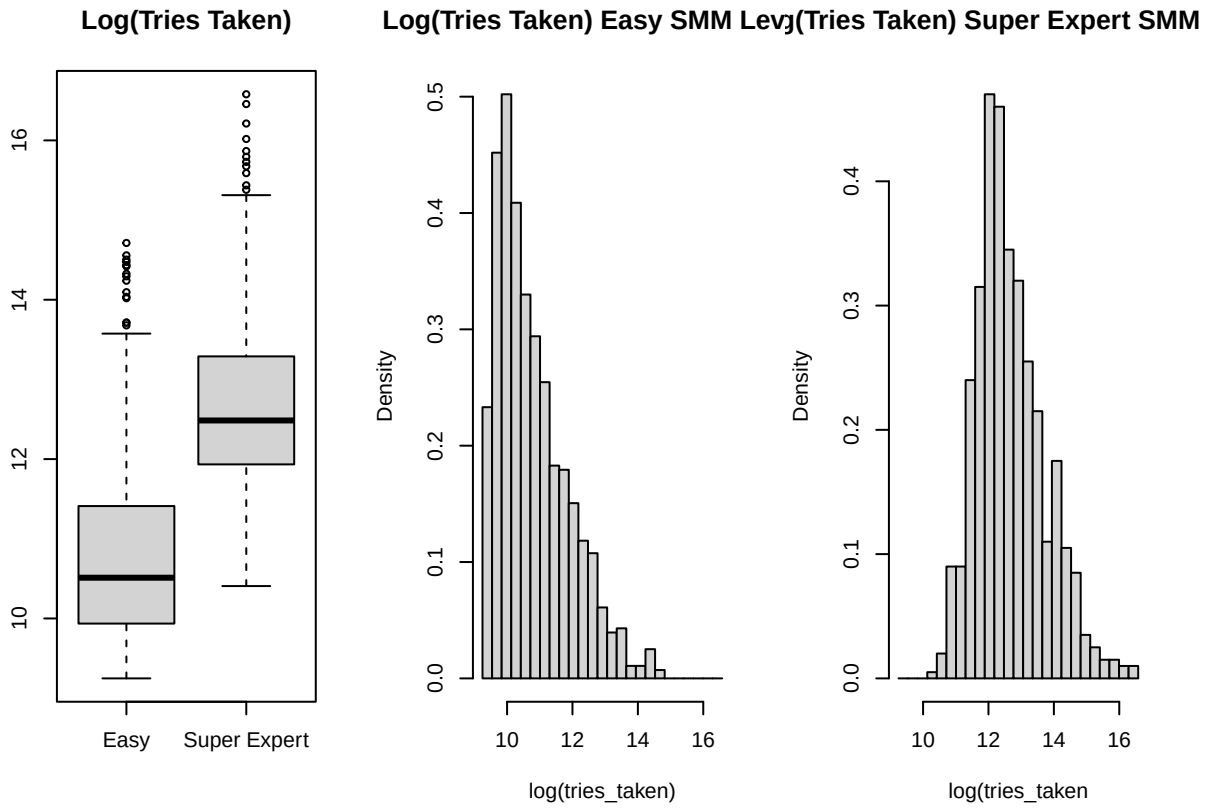
```
# ########
# REDACTED
# ########
```



The boxplot is a good representatino of this population as the transformed population is roughly unimodal.

We consider the log(number of tries taken) for each level while distinguishing by level difficulty. We construct a figure containing the boxplots and two histograms for the log(number of tries taken) by level difficulty using the same range for all the plots.

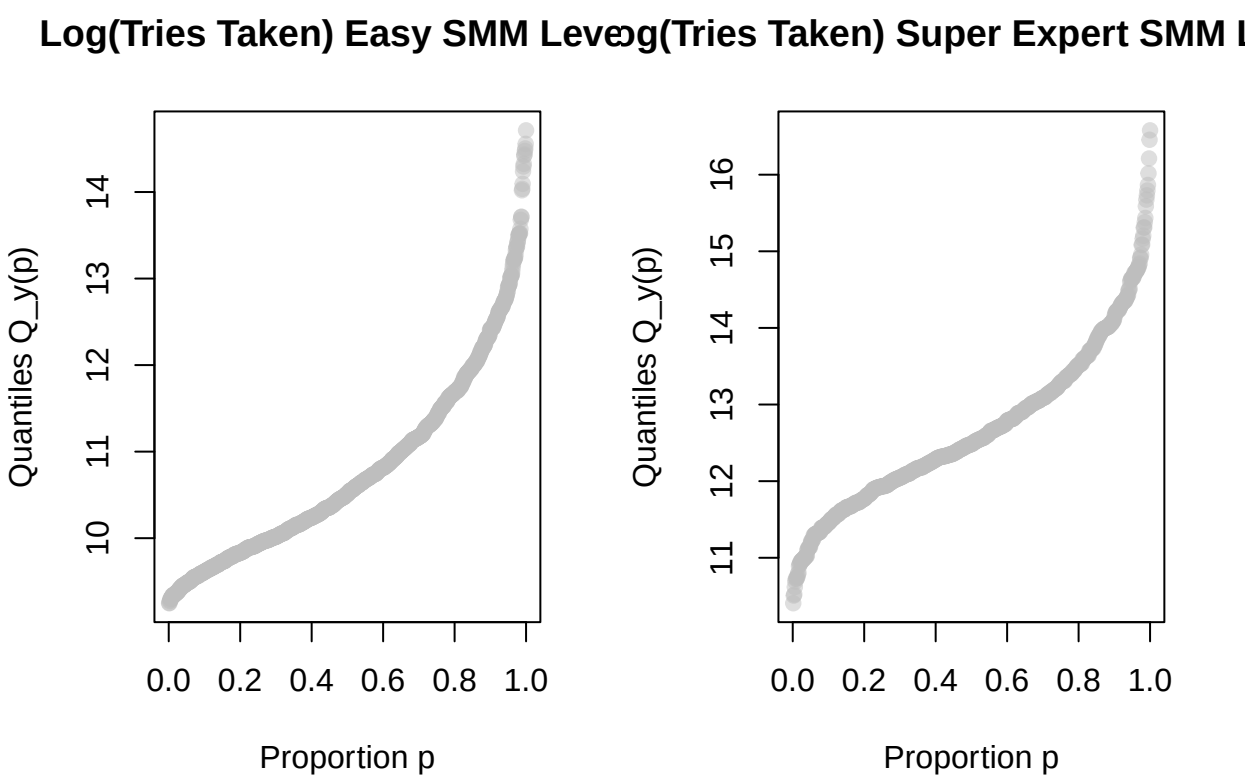
```
# ########
# REDACTED
# ########
```



Both the boxplot and histogram suggest that subset of levels which are super expert have a more symmetric distribution. Again the boxplots are good representations of these two populations as the transformed data is roughly unimodal.

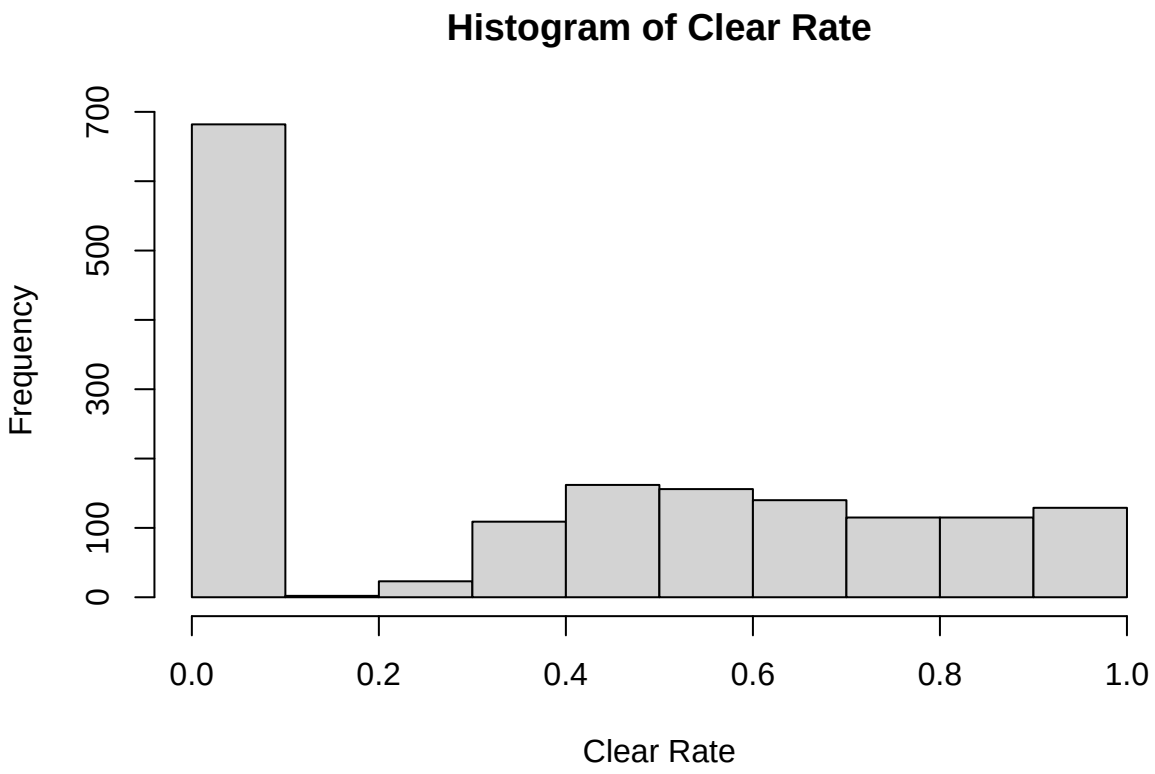
We construct a figure with one quantile plot for each level difficulty.

```
# ########
# REDACTED
# ########
```



The shape of the quantile plot for Easy levels suggests skewness, the shape of the quantile plot for the Super Expert levels suggests symmetry.

```
# ########
# REDACTED
# ########
```

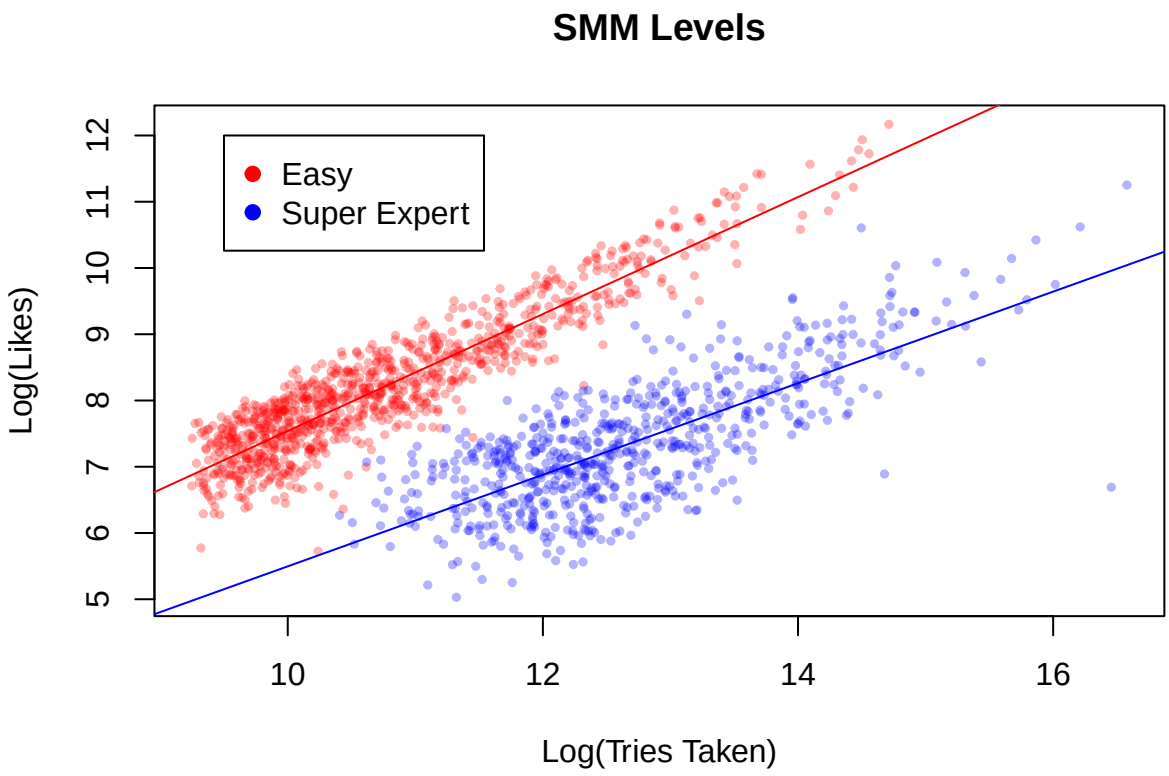


To summarize the clear_rate = tries_success/tries_taken we choose a histogram as the data is not unimodal. There is a large amount of the data in the low end and this is not evident from just the boxplot.

Influence on regression lines

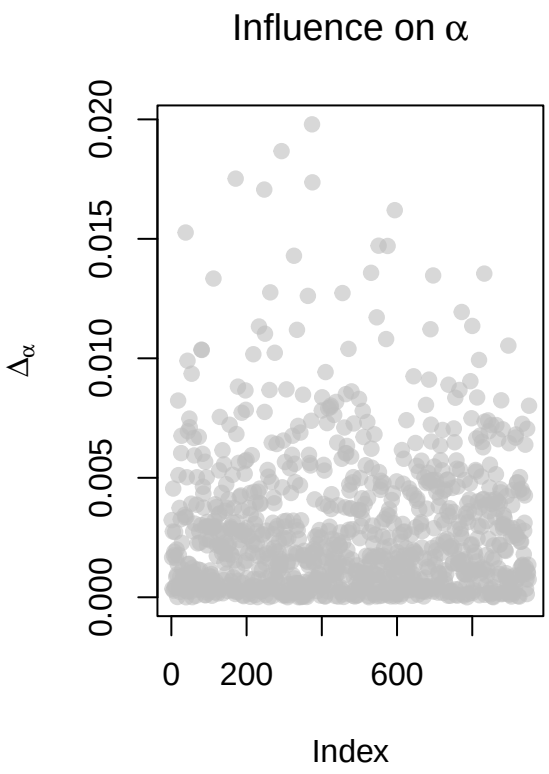
We create a scatterplot of log(tries_taken) and log(liked) while colouring the points by the level difficulty and add two regression lines (one for each difficulty type) by treating log(tries_taken)as the covariate and log(liked) as the response.

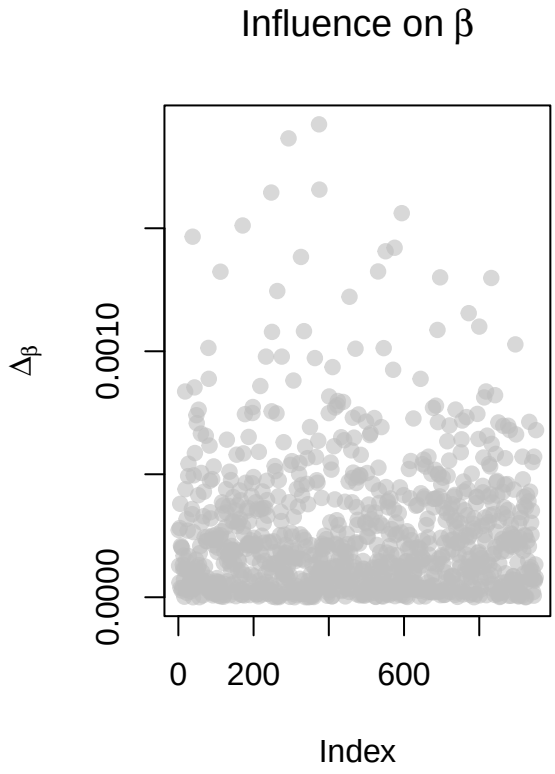
```
# ########
# REDACTED
# ########
```



For the Easy levels we calculate the influence of each level on the regression coefficient. In a figure, we plot the influence on the intercept and on the slope by level.

```
# ########
# REDACTED
# ########
```





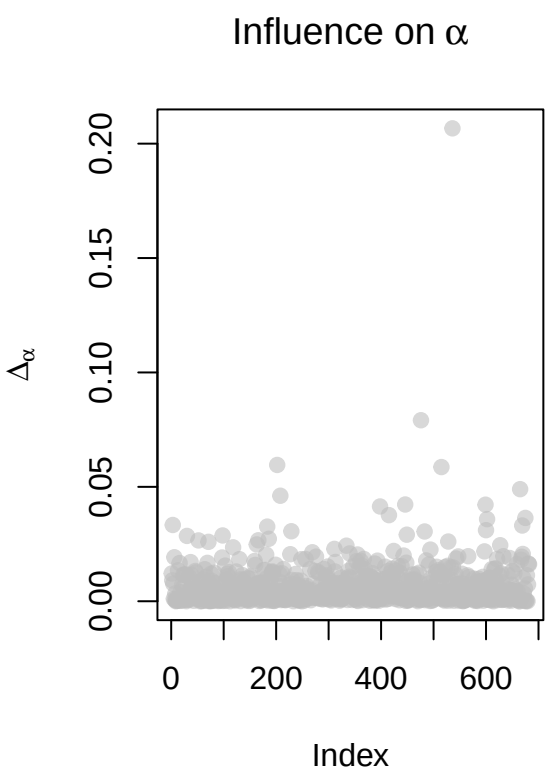
A table of the most influential points:

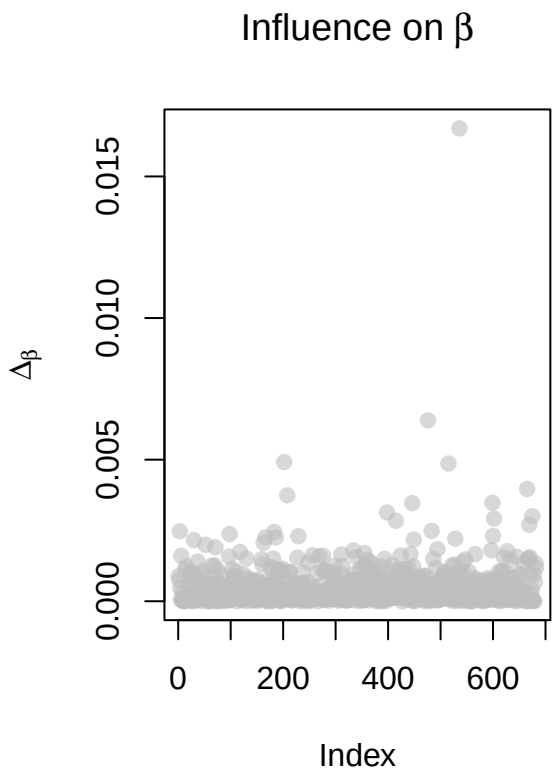
```
# ########
# REDACTED
# ########
```

	id	name	theme	created	difficulty	tag	liked	played	tries_taken	tries_success
284	2CE6-0000-00FB-3847	Try Not to Win	mario_bros3	2015-11-25	Easy	Gimmick	322	10028	11153	9772
411	423F-0000-0023-EEEB	! Gourmet Race	mario_world	2015-09-12	Easy	Music	91319	745799	871886	733969
481	4D77-0000-0115-E8F2	( )	mario_world	2015-12-12	Easy	Automatic	3733	65077	224318	61067
613	631E-0000-0128-B133		mario_bros_u	2015-12-23	Easy	NULL	13435	262805	554824	244405
615	635C-0000-0045-AF89	Automatic Mario Kart	mario_bros	2015-09-17	Easy	Automatic	192550	1594405	2452152	1445493
1011	A1FD-0000-0074-2C65	Pokémon Center Theme!	mario_bros	2015-09-27	Easy	Music	90488	695411	902485	653353

For the Super Expert levels, we calculate the influence of each level on the regression coefficient. In a figure, we plot the influence on the intercept and on the slope by level.

```
# ########
# REDACTED
# ########
```





A table of the most influential points:

```
# ########
# REDACTED
# ########
```

	id	name	theme	created	difficulty	tag	liked	played	tries_taken	tries_success
906	9227-0000-01BF-46CB	» « walk	mario_world	2016-02-07	Easy	Gimmick	2886	28897	65497	25756