Examining the association between an application’s panel score and its subsequent citations

Adrian Barnett

26 June 2018

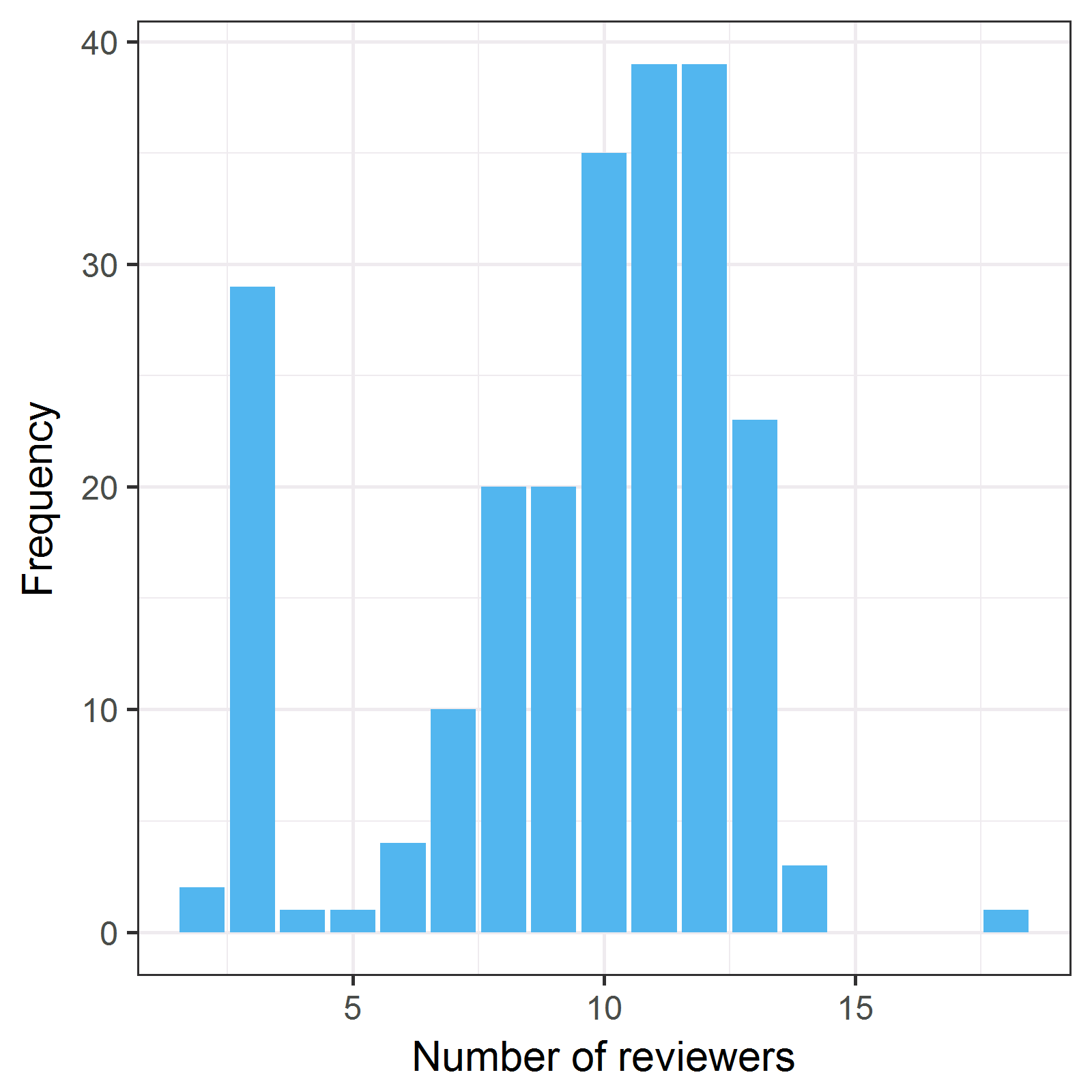
## Introduction

The data are 227 successful grant applications for the years 1999 to 2006. The outcome is citation number. The key predictor is the application score, which ranges from 1 (best) to 5 (worst). Funding decisions typically use the mean score to rank applications, but there may be value in also using the standard deviation in a score which reflects an application where there was some disagreement between panel members.

### Missing data

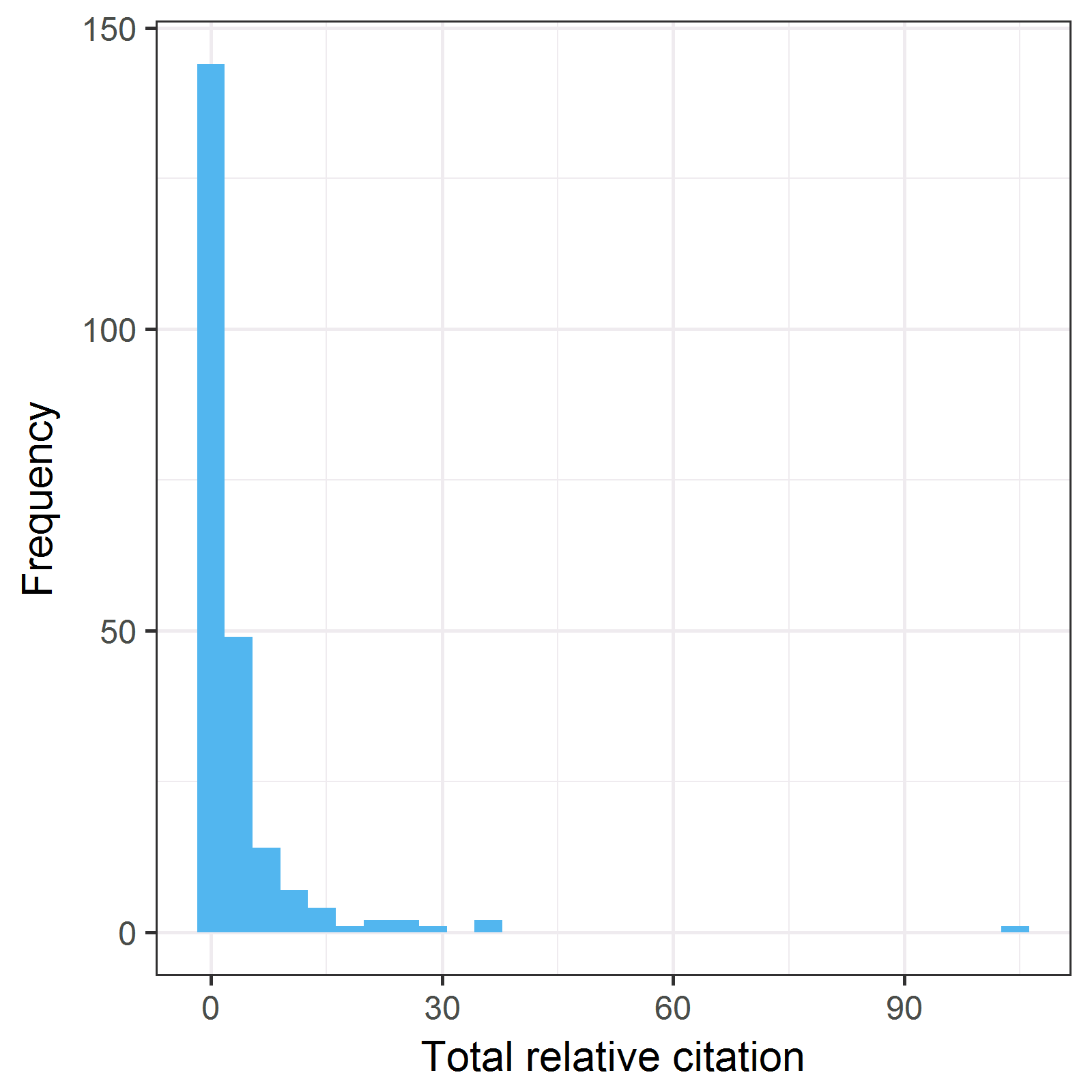
16 (7%) observations were missing the score standard deviation and 32 (14%) observations were missing the score range. These missing observations were imputed before any analysis.

### Bar plot of reviewer numbers



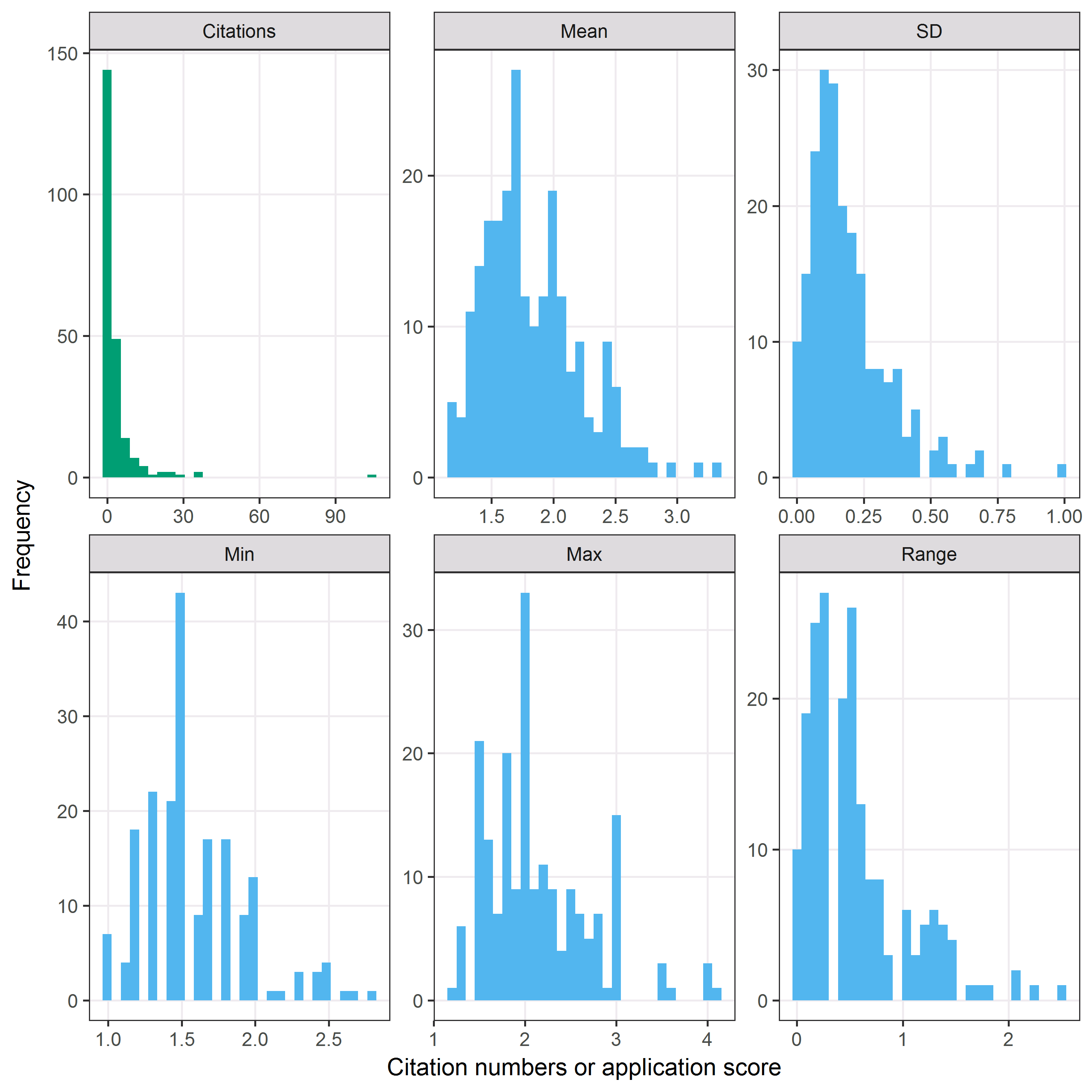
This is a bimodal distribution with a mode at 3 reviewers for the teleconference reviews and second mode at 11 to 12 for the larger panels.

#### Histogram of the primary outcome



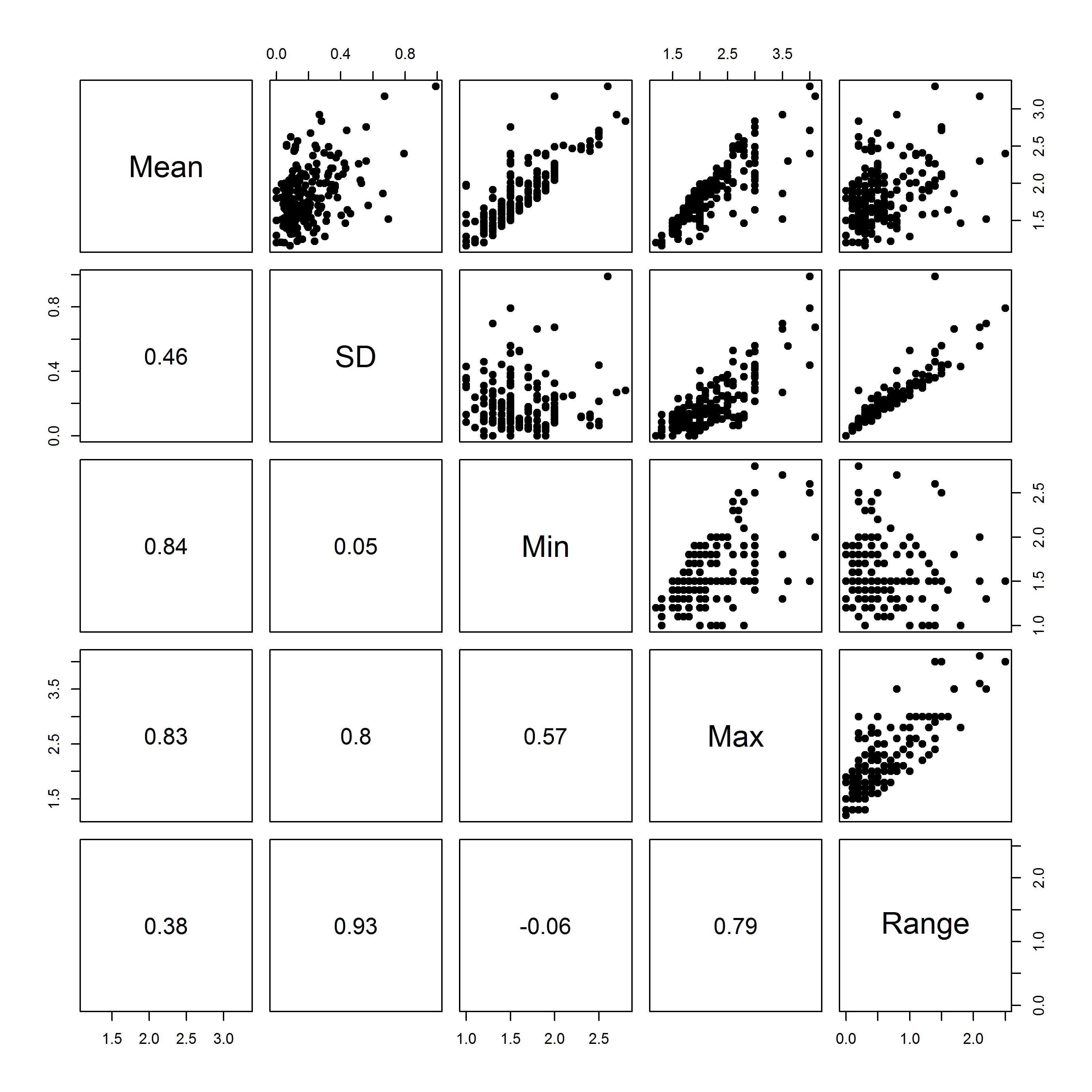
The main outcome is strongly positively skewed. There were 74 (33%) grants with zero citations.

#### Histograms of the application score statistics and citations



The histograms above show the mean, standard deviation, minimum, maximum and range (max minus min). The lower the mean score, the better the application did in peer review. The standard deviation and range have a positive skew.

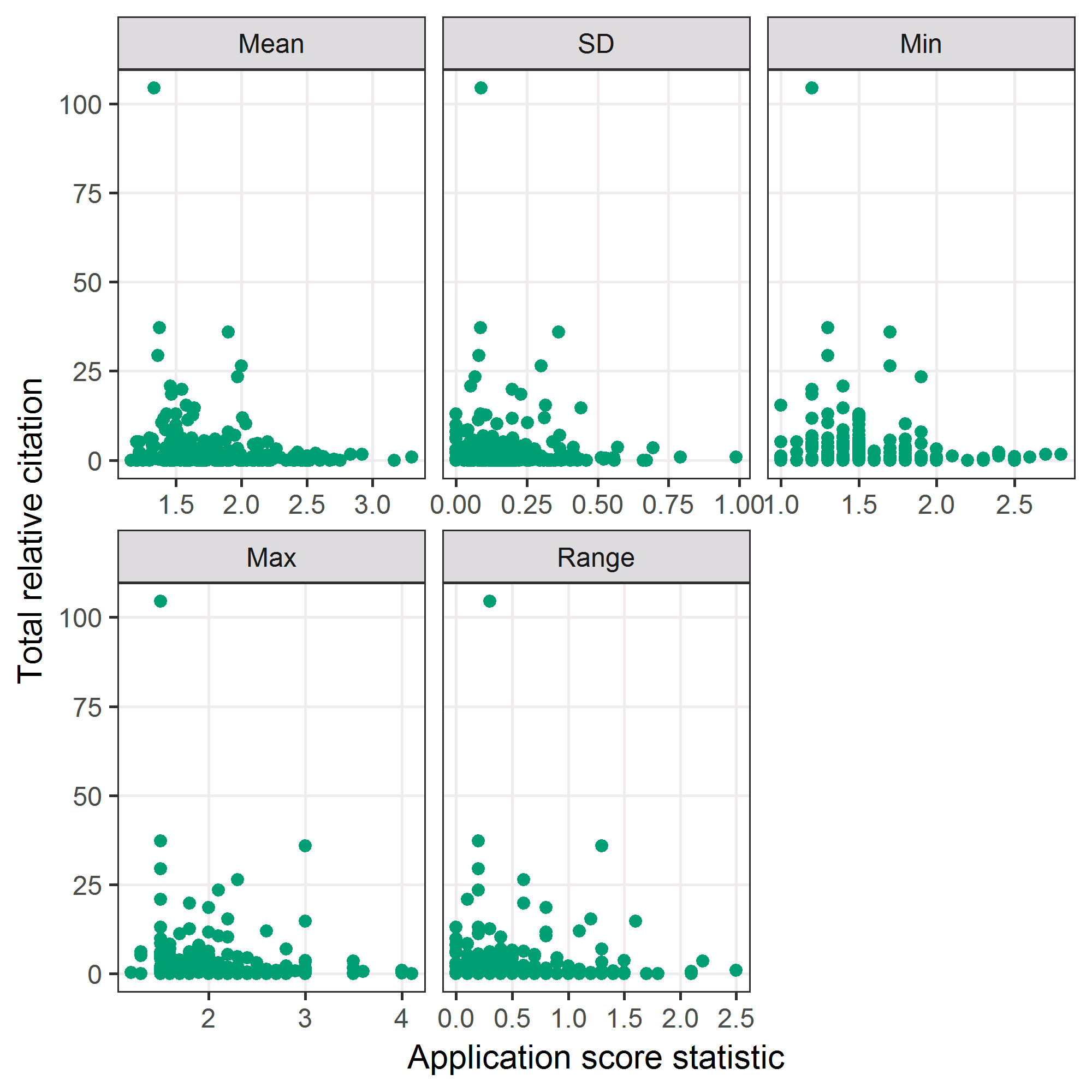
#### Scatter-plots and correlations between application score statistics



The numbers in the bottom-left half of the plot matrix are the Pearson correlations. We expect some positive correlation between a score’s mean and standard deviation (0.46). There is a relatively strong correlation between the standard deviation and maximum (0.80), but not between the standard deviation and minimum (0.05). This suggests the largest disagreement is where at least one panel member has given a poor score. Applications where there was one dissenting panel member with a good score were probably not funded as their mean score would not be competitive, and hence are not in this sample.

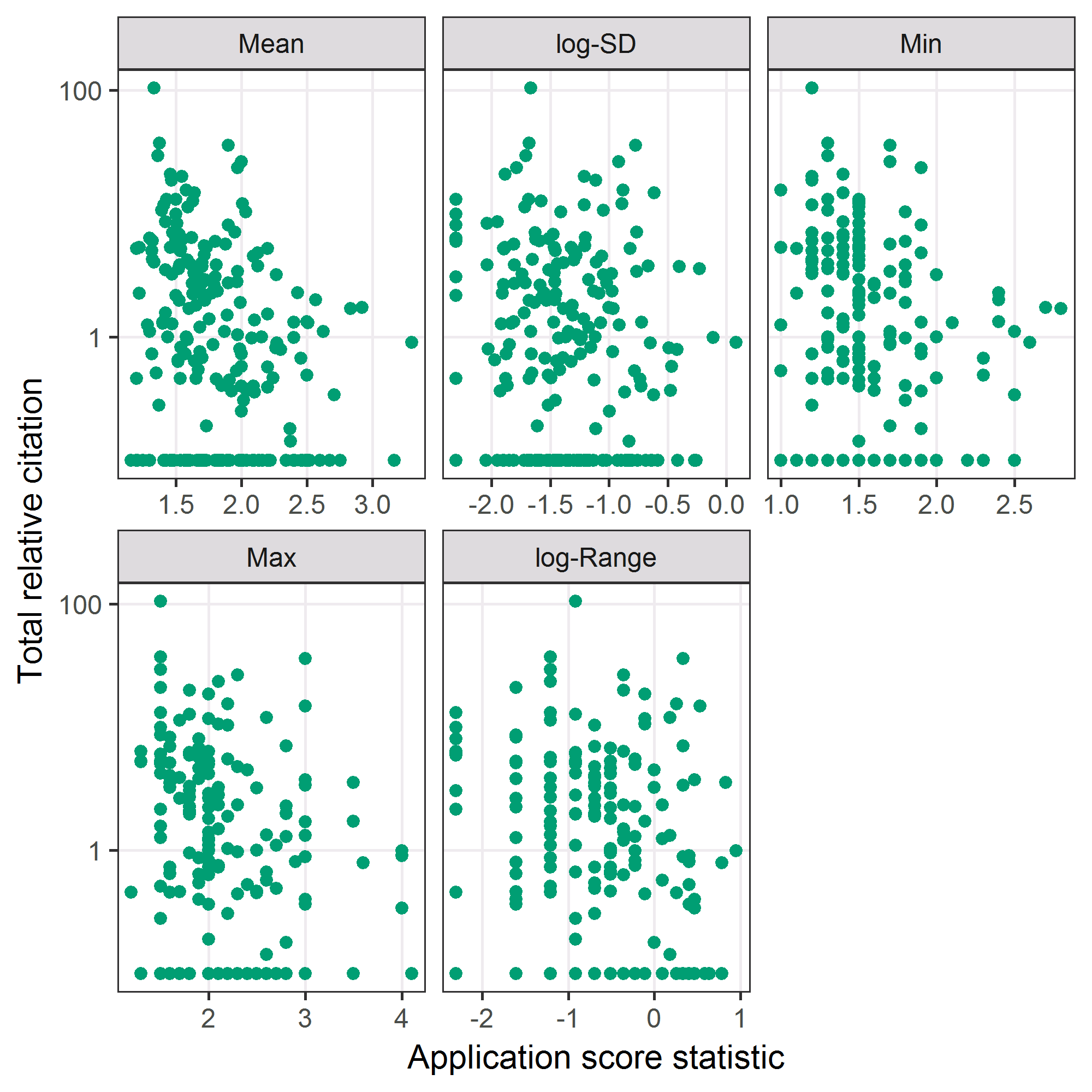
There is a strong positive correlation between the standard deviation and the range (0.93), and both are measures of panel disagreement.

### Scatter plots of citations against score statistics



The top-left panel showing the mean is a repeat of Figure 2 from the original PLOS ONE paper.

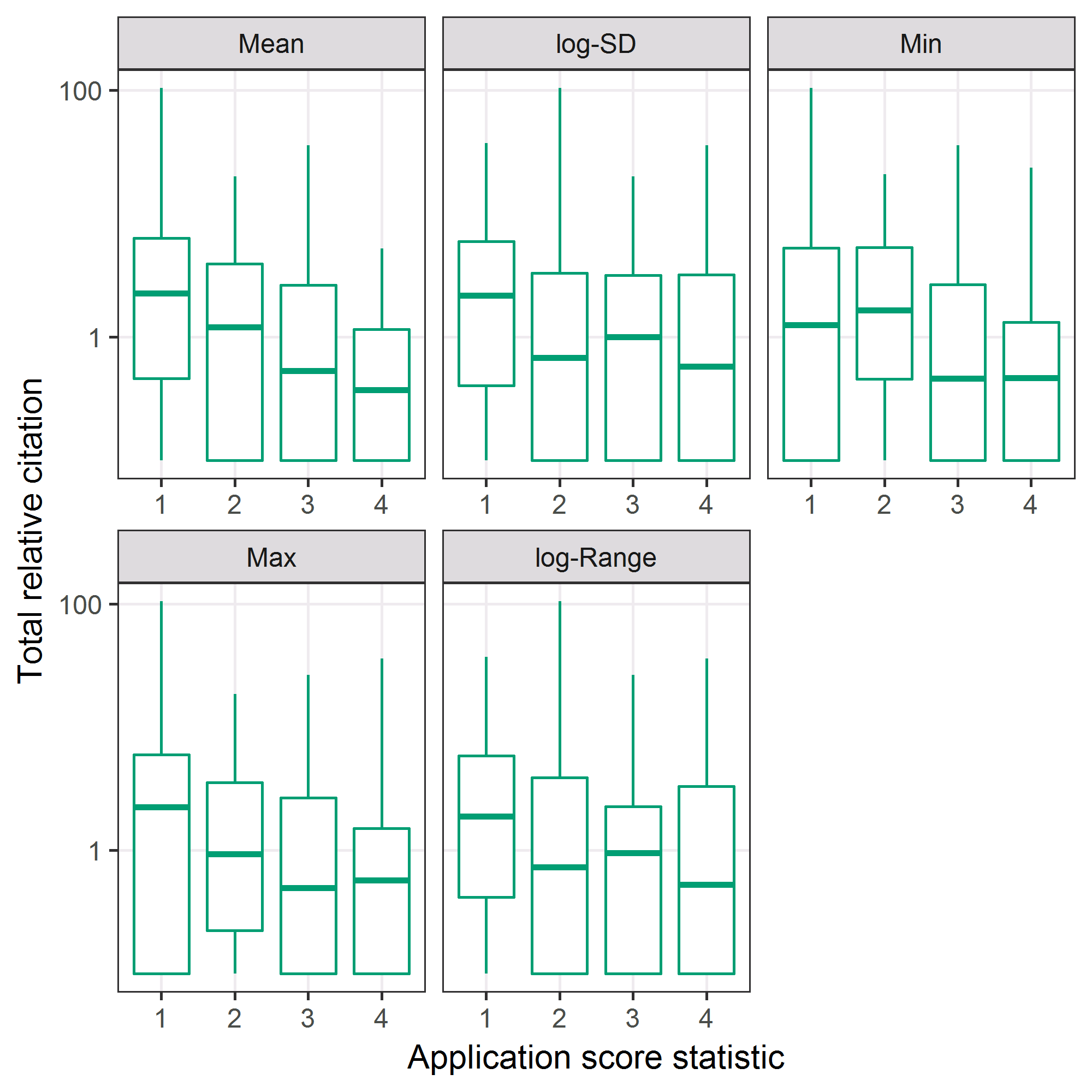
### Scatter plots of log-citations against score statistics



The above scatter plot uses log-transformed citations in an attempt to show a clearer association with the score statistics. Some association between mean score and citations is visible, with a generally downward pattern in citations for increasing score.

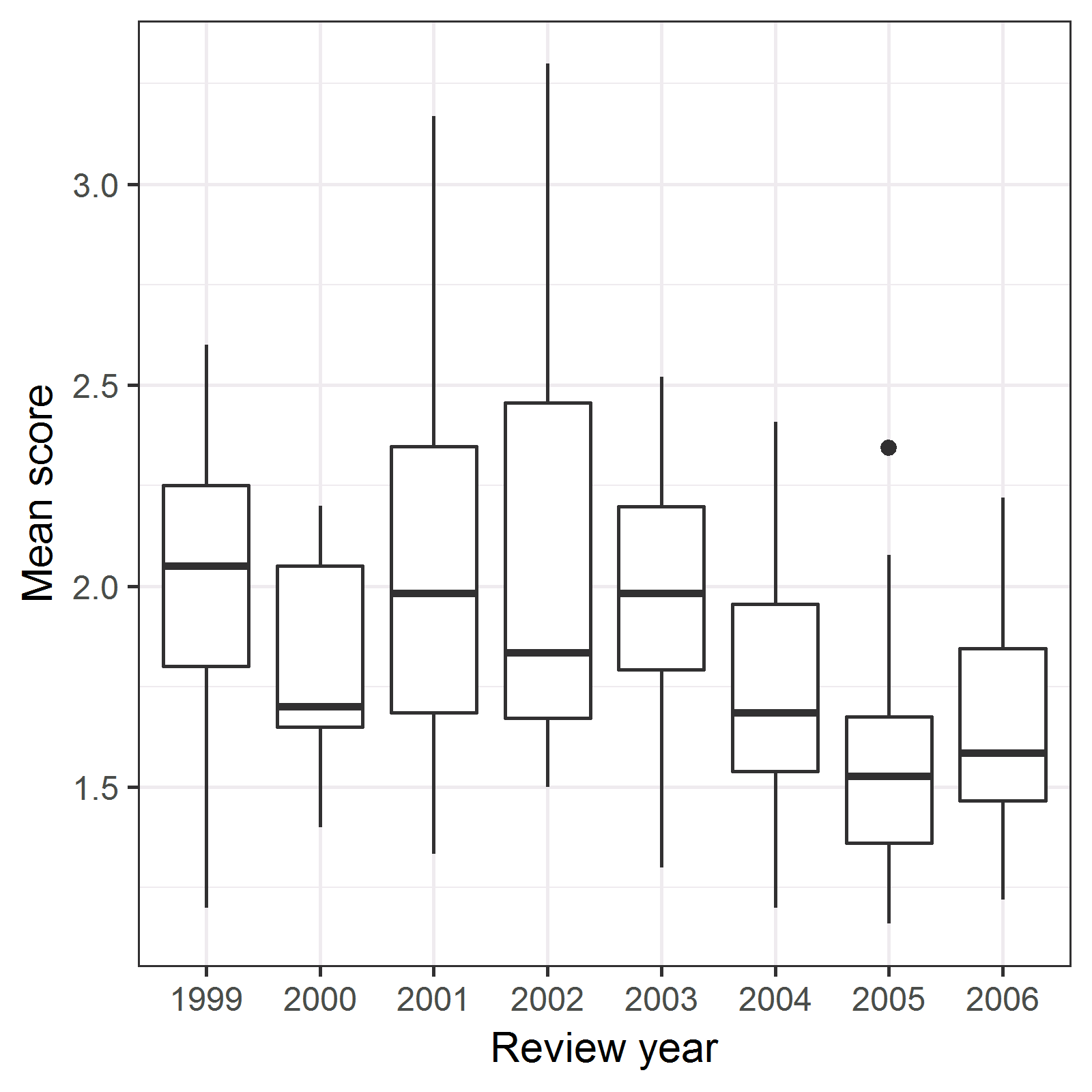
#### Checking for constant variance

The score statistics have been grouped into quintiles to check the variance.

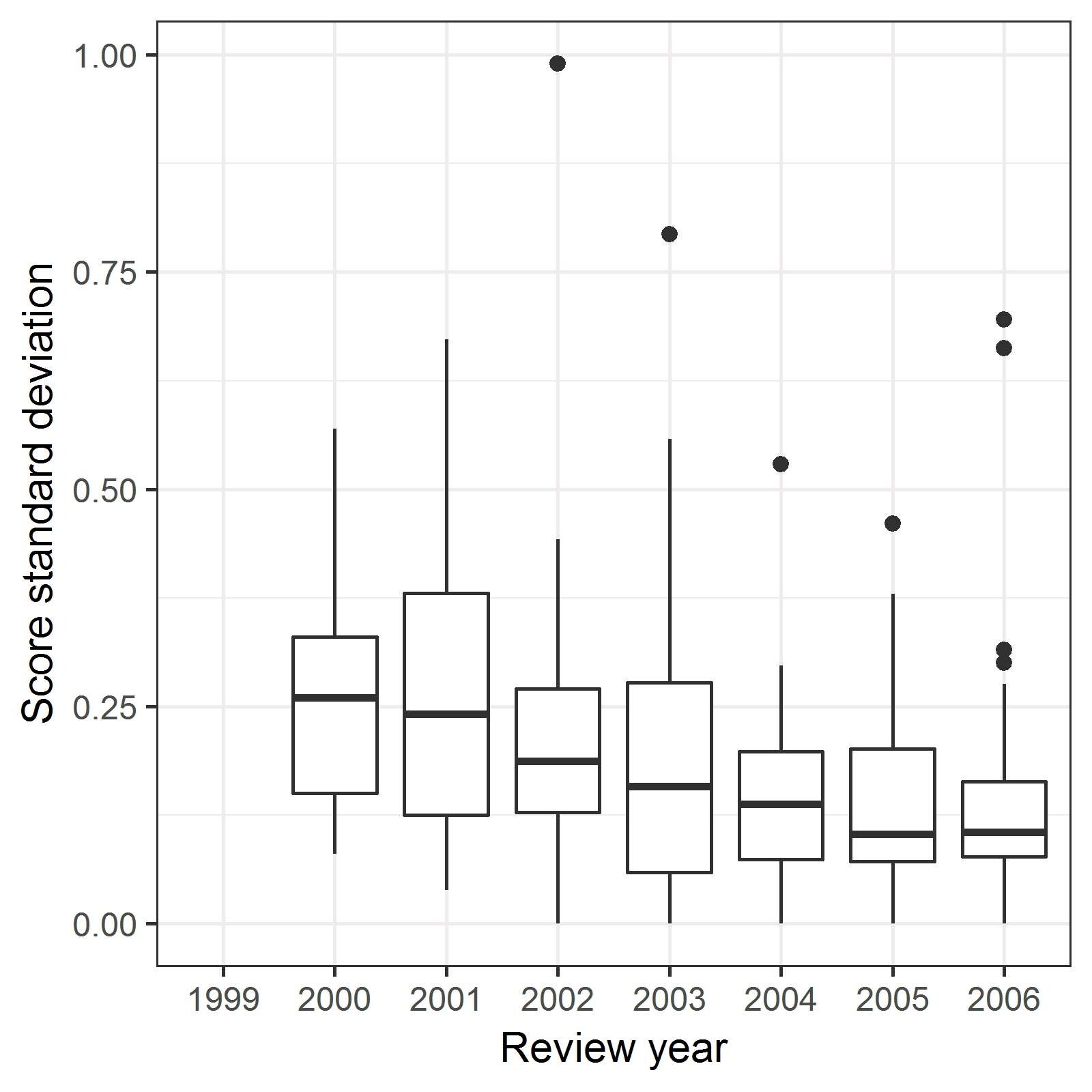


#### Changes in application scores over time

##### Mean scores



##### Score standard deviation



There has been some reduction in the mean and standard deviation over time. Hence it would be wise for our regression models to adjust for review year. This change was highlighted in the original PLOS paper and the plot of the mean score over time was also in the paper.

## Regression models

The total relative citations were modelled using a multiple regression model. The citations were first log-transformed because of their positive skew. We added a small positive constant of 0.1 prior to using the log-transform because some citations were zero. The predictions were back-transformed to show the estimates on the original citations scale.

We used fractional polynomials to examine a range of non-linear associations between the scores and citations.

Using equations the regression model was:

The log-transformed citations (*Y*) were modeled using a Normal distribution. The mean () had a constant and the application score statistics (*X*) which were first transformed by the fractional polynomial function. The random intercept () adjusted for a potential within-panel correlation in citations.

The 16 missing score standard deviations and 32 missing ranges were accounted for using multiple imputation.

### Model with just mean score

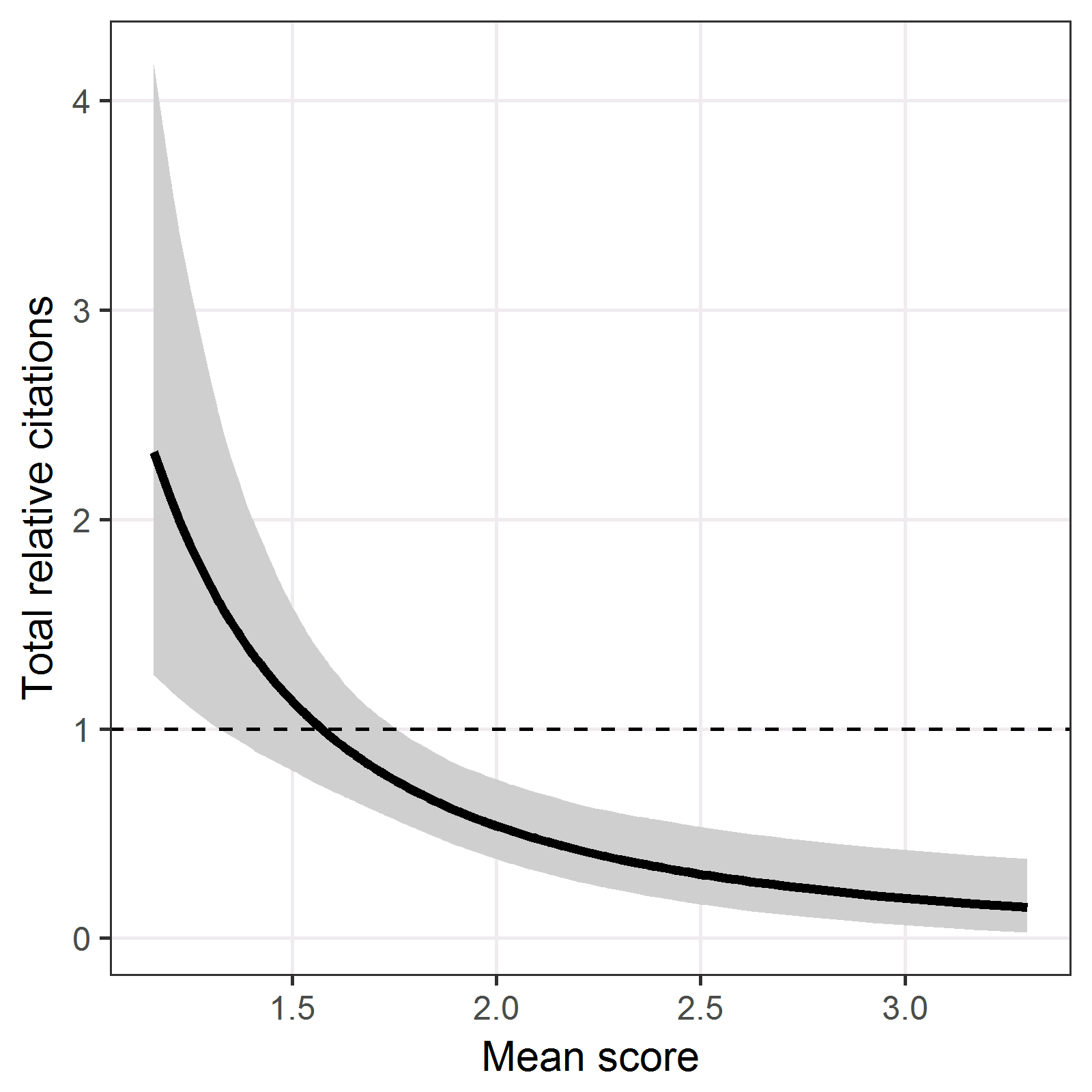
As a first step we fit just the mean score to confirm the findings from the previous paper.

#### Model fit - deviance

|  |  |
| --- | --- |
| Fractional polynomial | deviance |
| -2 | 886.9 |
| -1 | 885 |
| -0.5 | 883.6 |
| 0 | 885.4 |
| 0.5 | 884.6 |
| 1 | 887 |
| 2 | 891.1 |
| 3 | 895.5 |

The table shows the lowest deviance is for a fractional polynomial of -0.5, although the difference between the three best models is small.

#### Predictions from best model



The plot shows the predicted relative citations using the best fractional polynomial model. The solid line is the mean and the grey area is a 95% confidence interval.

#### Table of parameter estimates

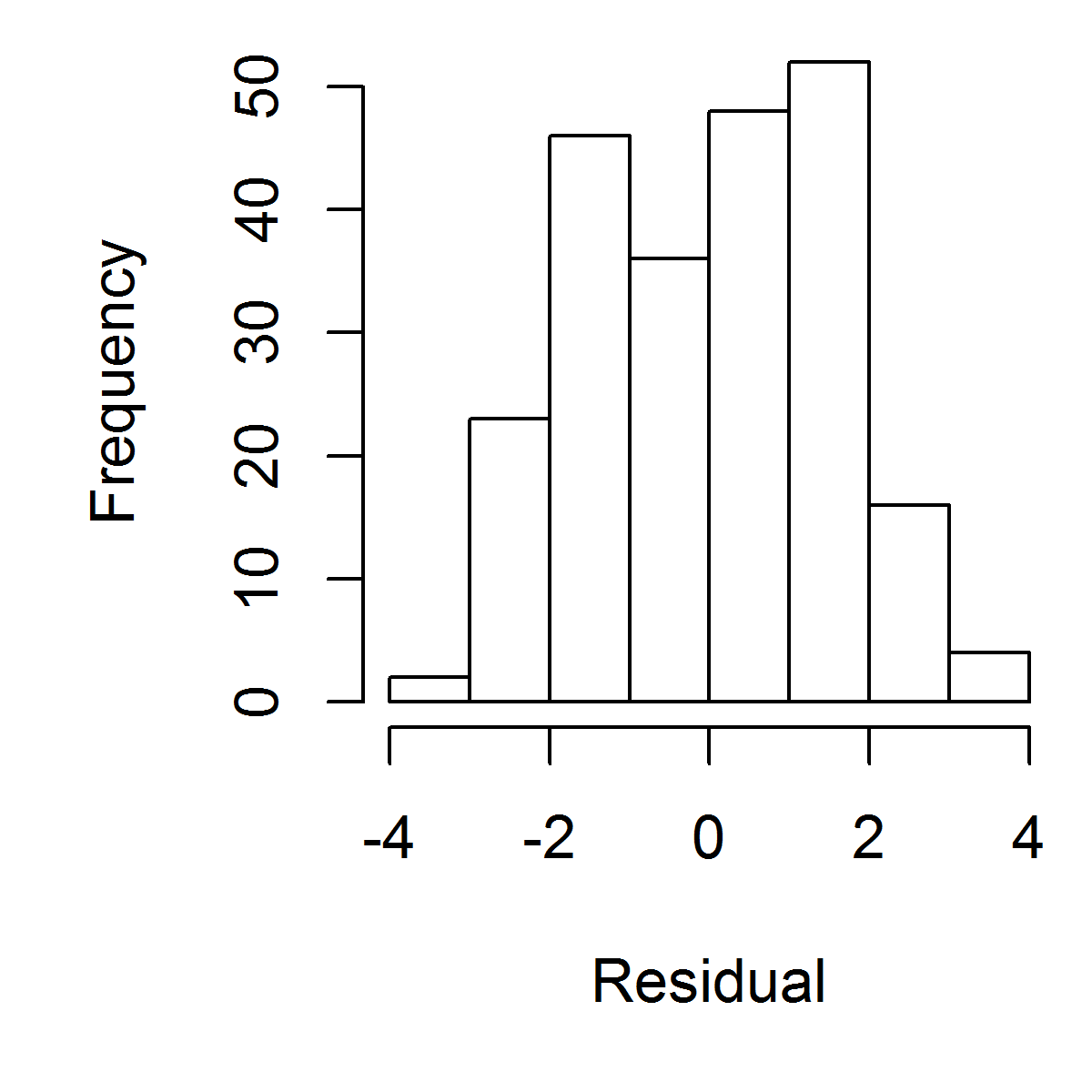
========================================  
 Dependent variable:   
 --------------------  
 log(`Web TRC` + 0.1)  
----------------------------------------  
Mean score, f(-0.5) 6.1 (3.1 to 9.0)   
Intercept -4.7 (-7.0 to -2.5)   
----------------------------------------  
Observations 227   
Log Likelihood -441.8   
========================================  
Note:

The table shows the mean and 95% confidence interval for the parameter estimates (%%). There was a strongly statistically significant effect of the mean as the 95% confidence interval does not contain zero. However, it is difficult to interpret these estimates because of the fractional polynomial transformation. Hence it is better to interpret the effect using the previous figure.

The within-panel correlation was 0.11, signifying a small positive correlation in the citations of applications assessed by the same panel.

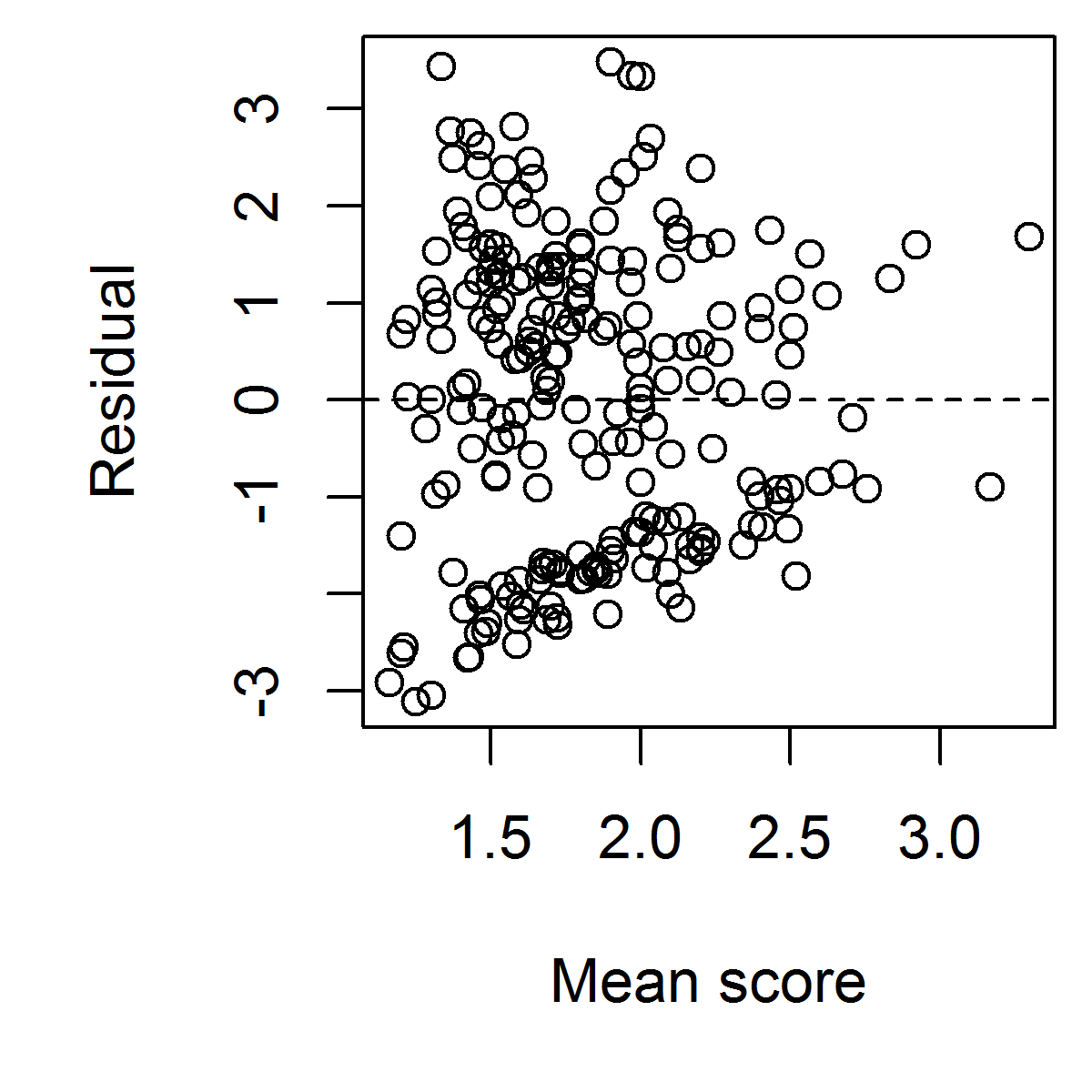
#### Model checking

##### Histogram of residuals

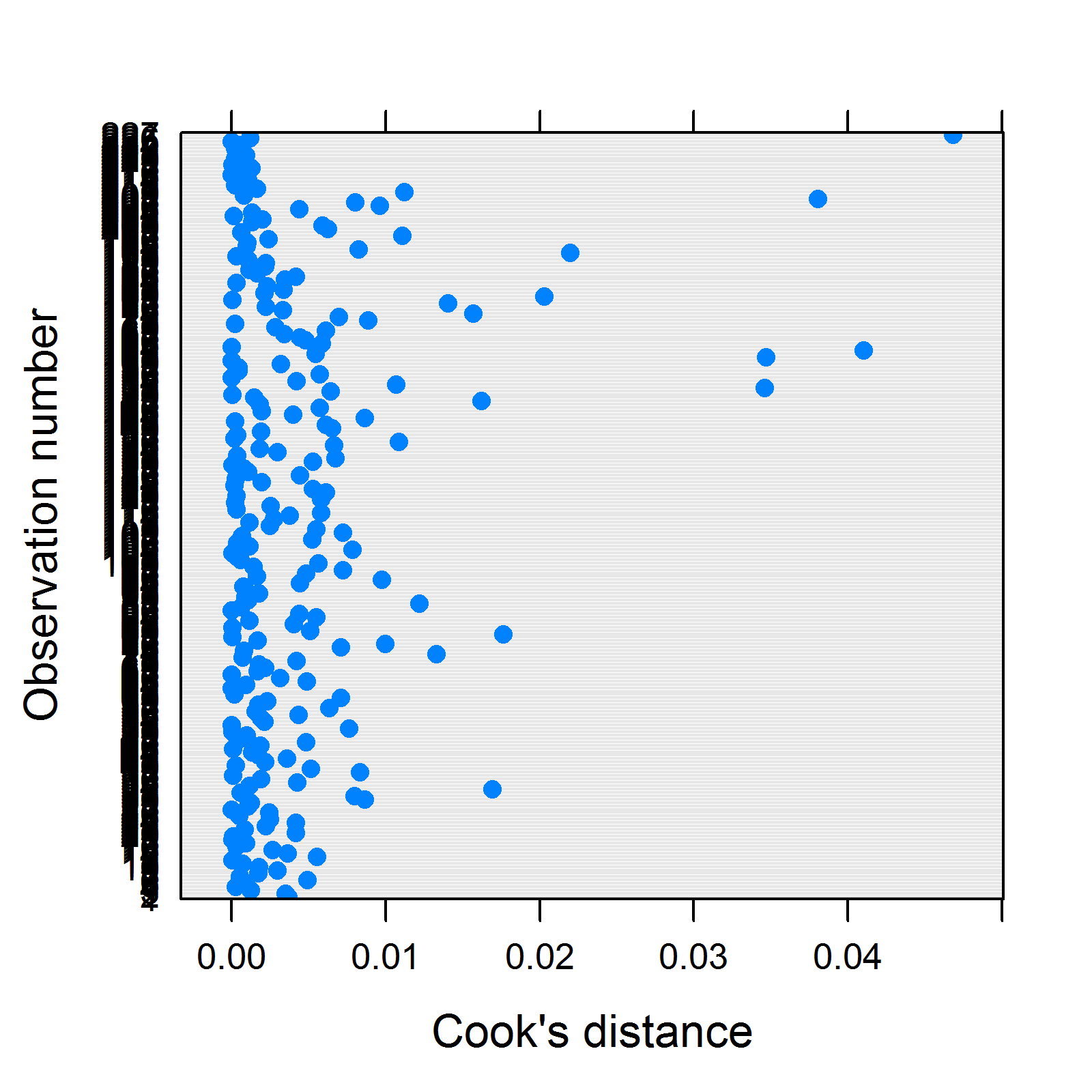


There are no clear outliers in the residuals.

##### Scatter plot of residuals against mean

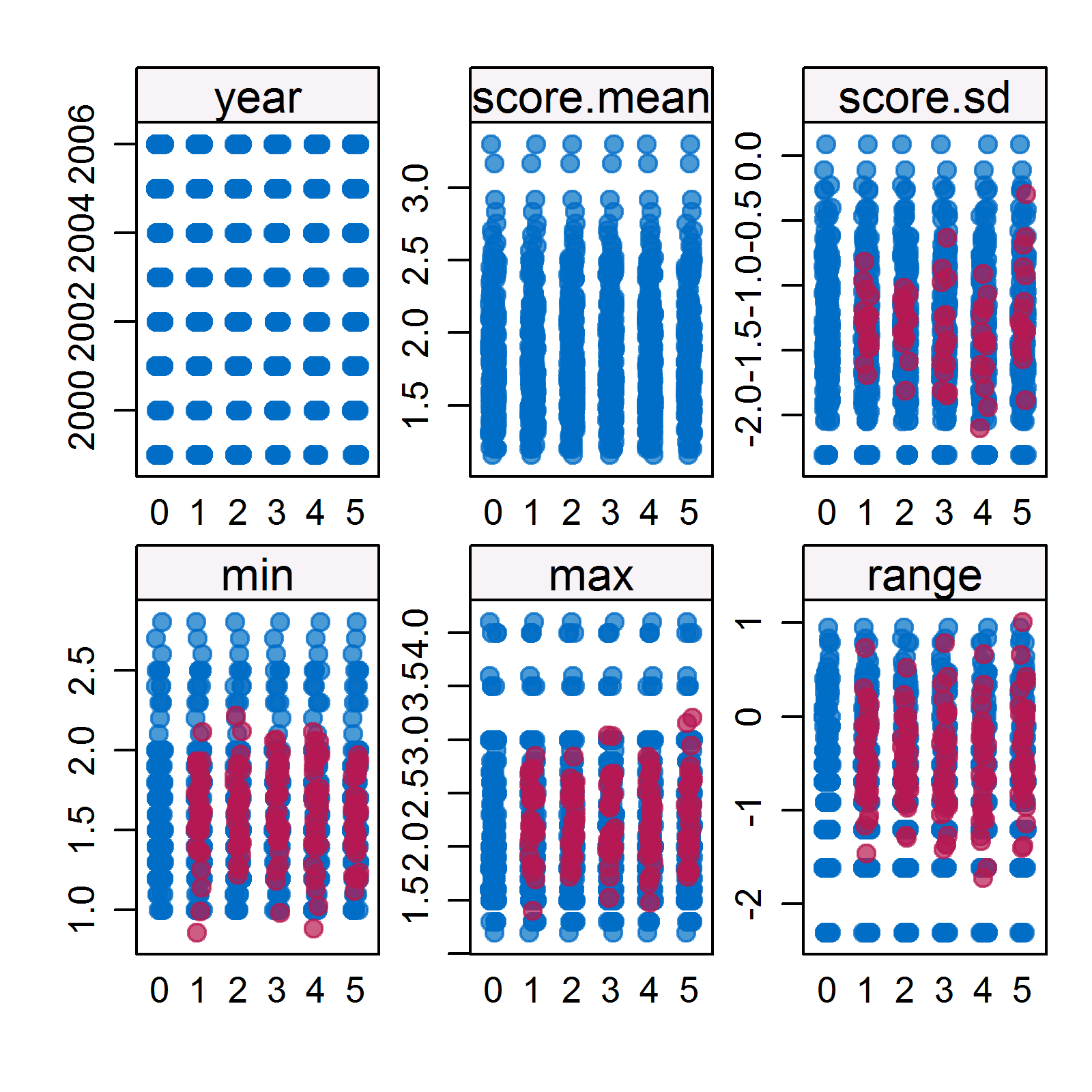


##### Cook’s distance for finding influential observations



## Imputing missing standard deviations and ranges

We used regression imputation to impute the missing standard deviations and ranges based on the mean, minimum and maximum score. We use the *mice* library in R. We log-transformed the standard deviation and range because of their positive skew. We created five imputed data sets. The plot shows the observed (blue) and imputed (red) observations.



### Model with mean score and score standard deviation

#### Model fit - Top five deviances (FP = Fractional Polynomial)

|  |  |  |  |
| --- | --- | --- | --- |
| FP1 | FP2 | FP3 | Deviance.mean |
| -0.5 | 3 | -0.5 | 878 |
| -0.5 | 0.5 | -0.5 | 878.4 |
| -0.5 | 3 | -1 | 878.4 |
| -0.5 | 2 | -0.5 | 878.4 |
| 0.5 | 3 | -0.5 | 878.6 |

AS before the best fractional polynomial for the mean was -0.5.

#### Predictions from best model

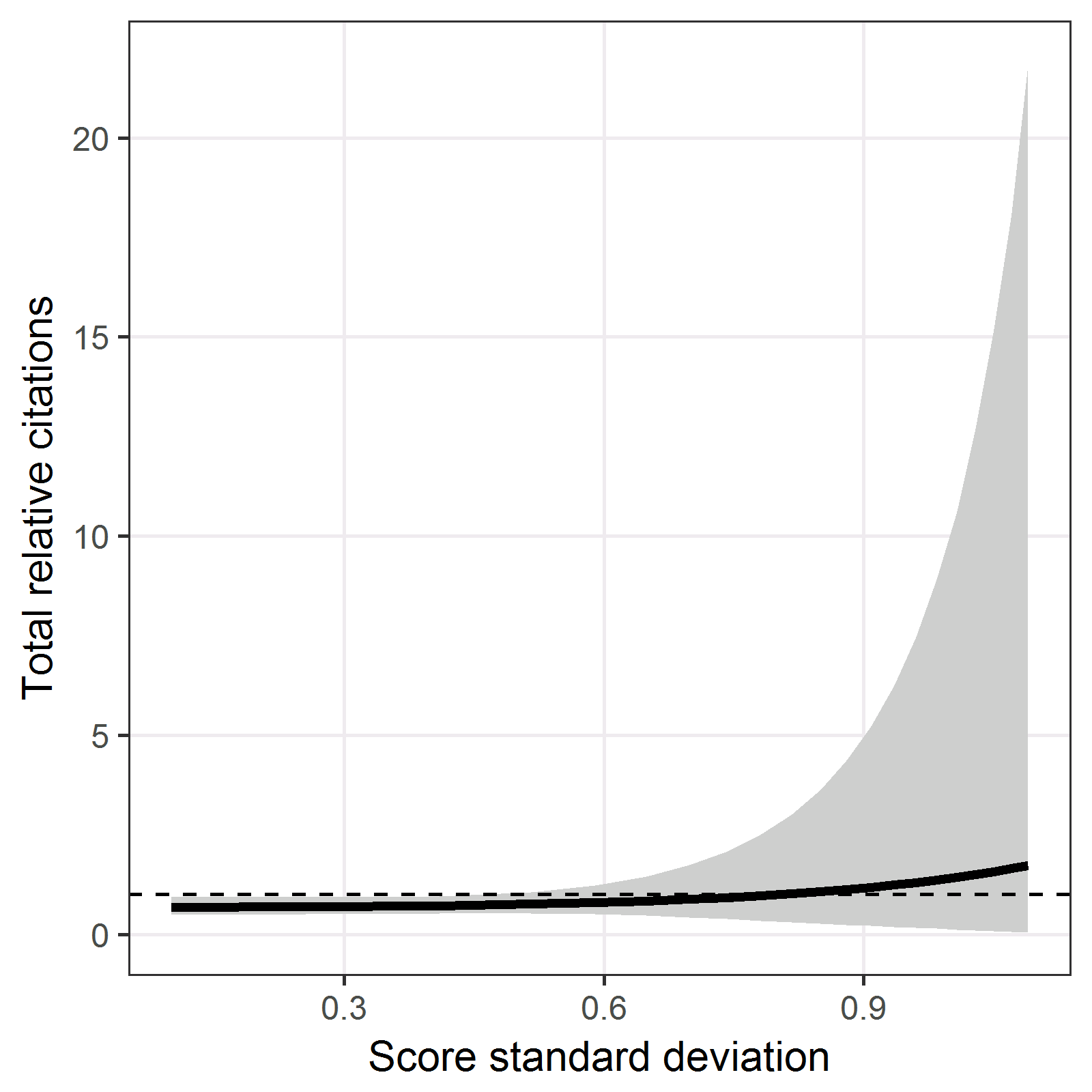
##### Predictions for mean score

Multiple imputation results:  
 MIcombine.default(betas, vars)  
 results se (lower upper) missInfo  
(Intercept) -4.0523014 1.4479822 -6.890295 -1.2143083 0 %  
score.tran 5.8722034 1.6550863 2.628294 9.1161130 0 %  
score.tran2 0.6504743 1.0086011 -1.326366 2.6273149 1 %  
score.tran3 -1.2168051 0.9297673 -3.039115 0.6055053 0 %



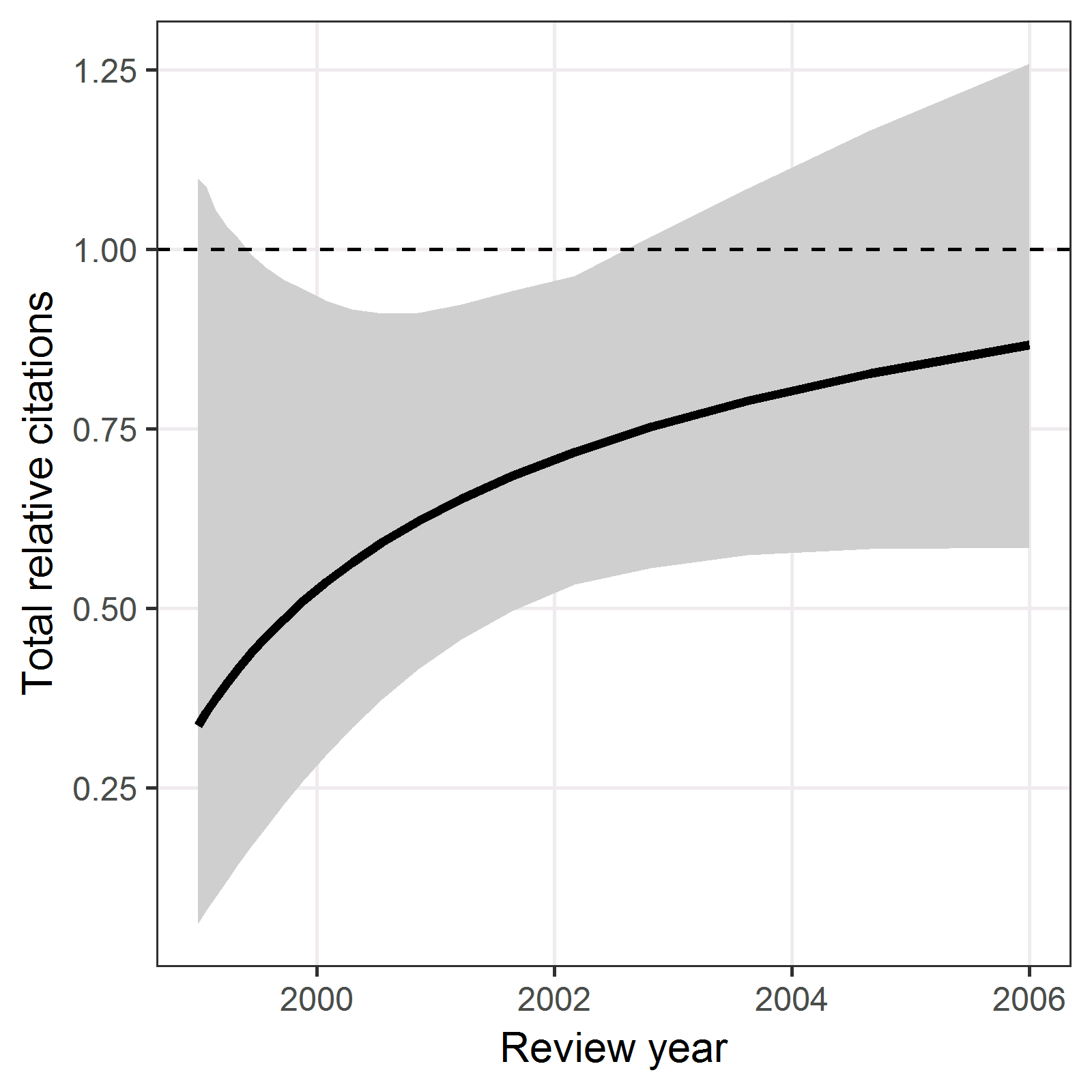
The mean shows a reduction in total citations for higher scores. The reduction for scores between 1 to 2 is steeper than the reduction for scores between 2 to 3.

##### Predictions for standard deviation



The mean line is flat, showing no clear association between the standard deviation and total citations. The very wide confidence interval indicates great uncertainty for high standard deviations.

##### Predictions for review year



There was a relatively small increase in citations by review year.

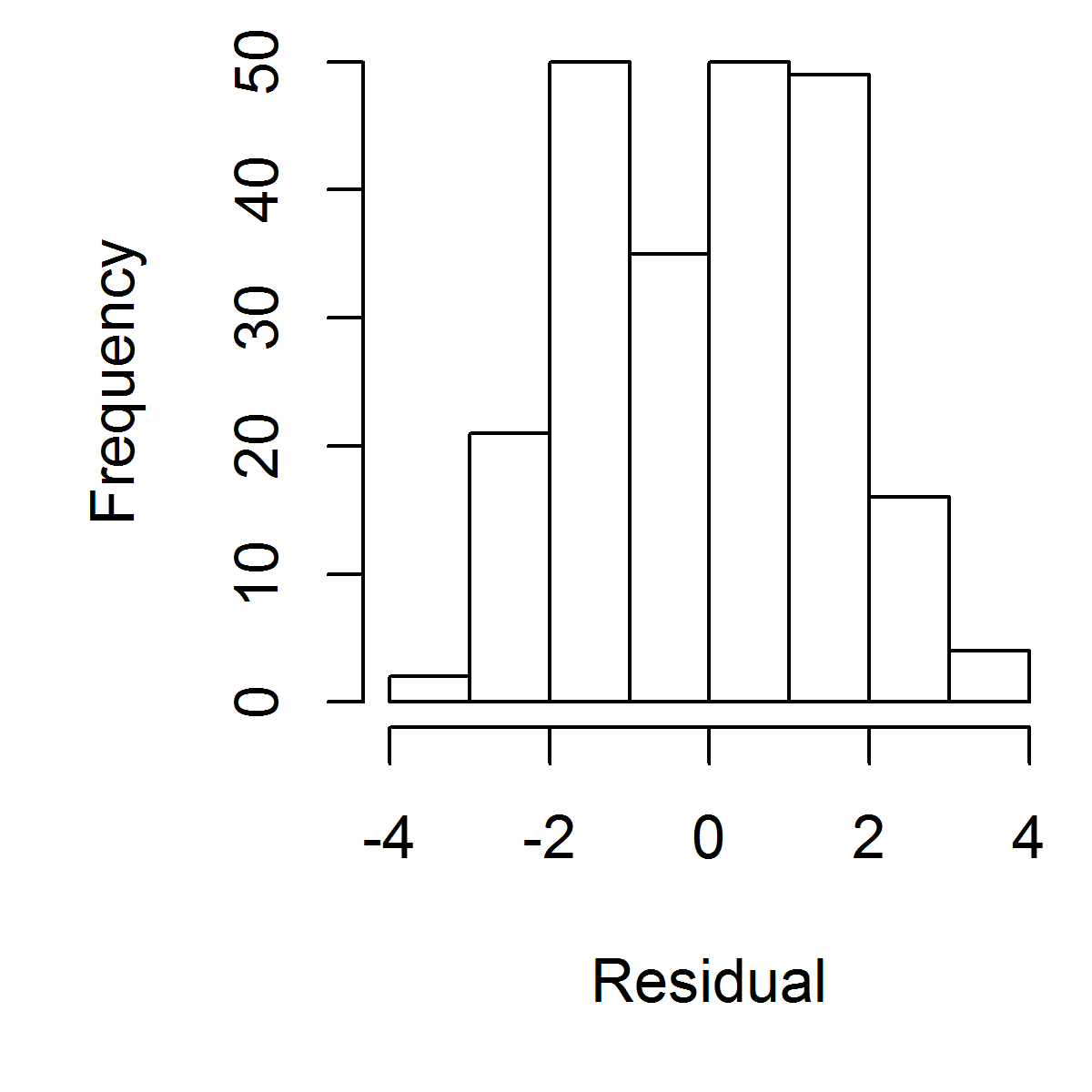
#### Table of parameter estimates

=========================================  
 Dependent variable:   
 --------------------  
 log(`Web TRC` + 0.1)  
-----------------------------------------  
Mean score, f(-0.5) 5.9 (2.6 to 9.1)   
Score SD, f(3) 0.7 (-1.3 to 2.6)   
Review year, f(-0.5) -1.2 (-3.0 to 0.6)   
Intercept -4.1 (-6.9 to -1.2)   
-----------------------------------------  
Observations 227   
Log Likelihood -439.0   
=========================================  
Note:

Only the mean score had a statistically significant effect on citations. As before it is better to use the plots rather than the parameter estimates to interpret the estimated effects. The “f()” in the variable label is the best fractional polynomial transformation.

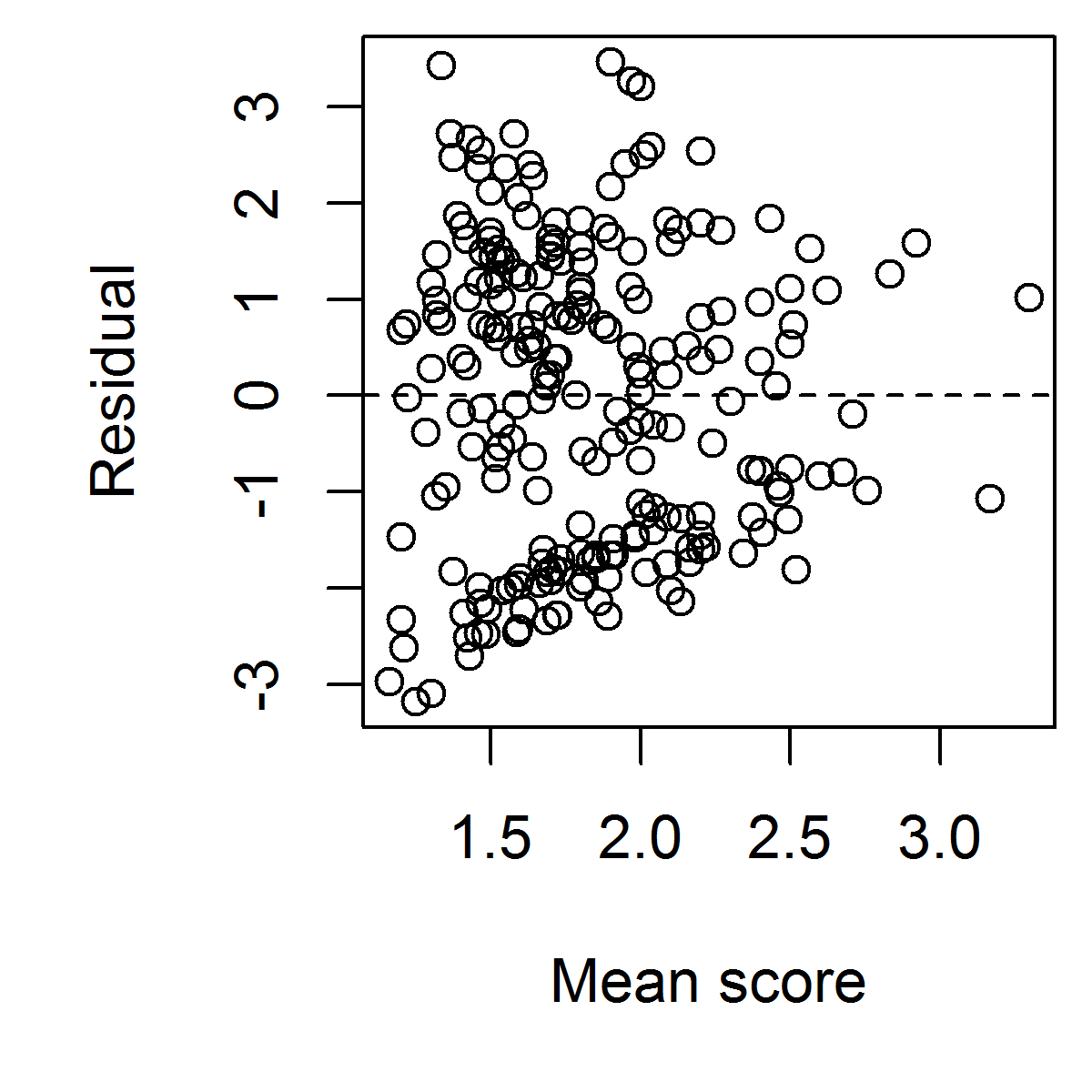
#### Model checking

##### Histogram of residuals

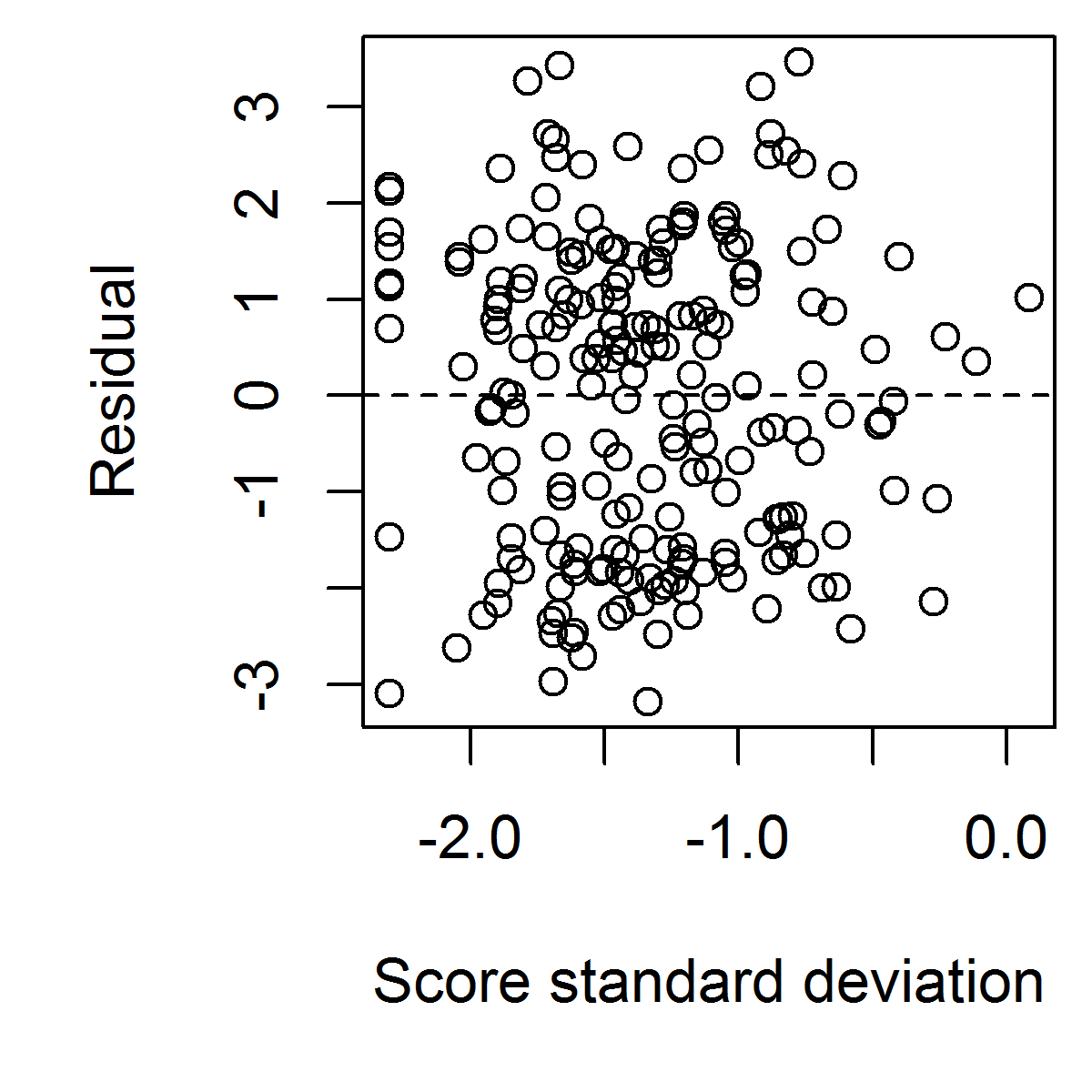


There are no clear outliers in the residuals.

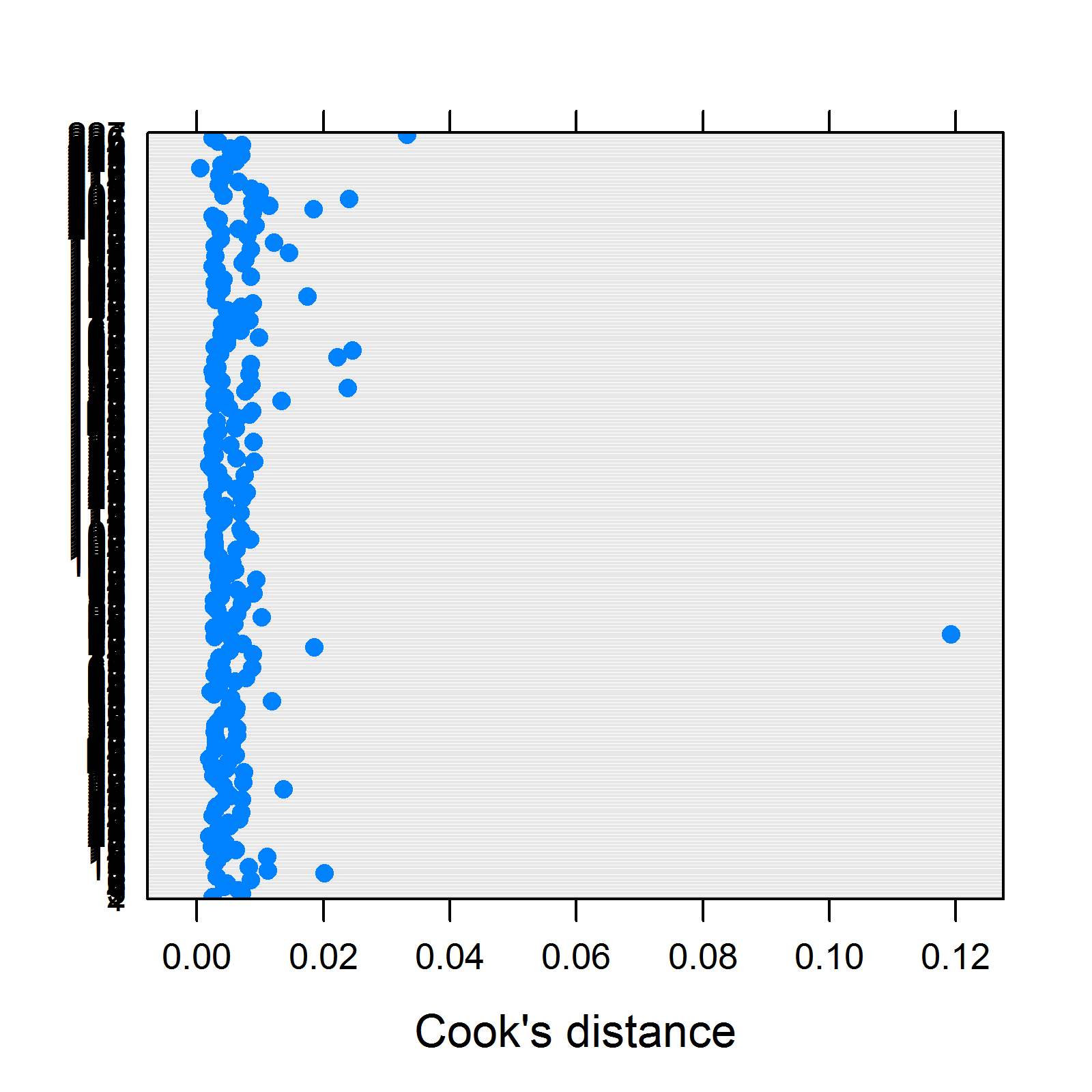
##### Scatter plot of residuals against mean



##### Scatter plot of residuals against log standard deviation



##### Cook’s distance for finding influential observations



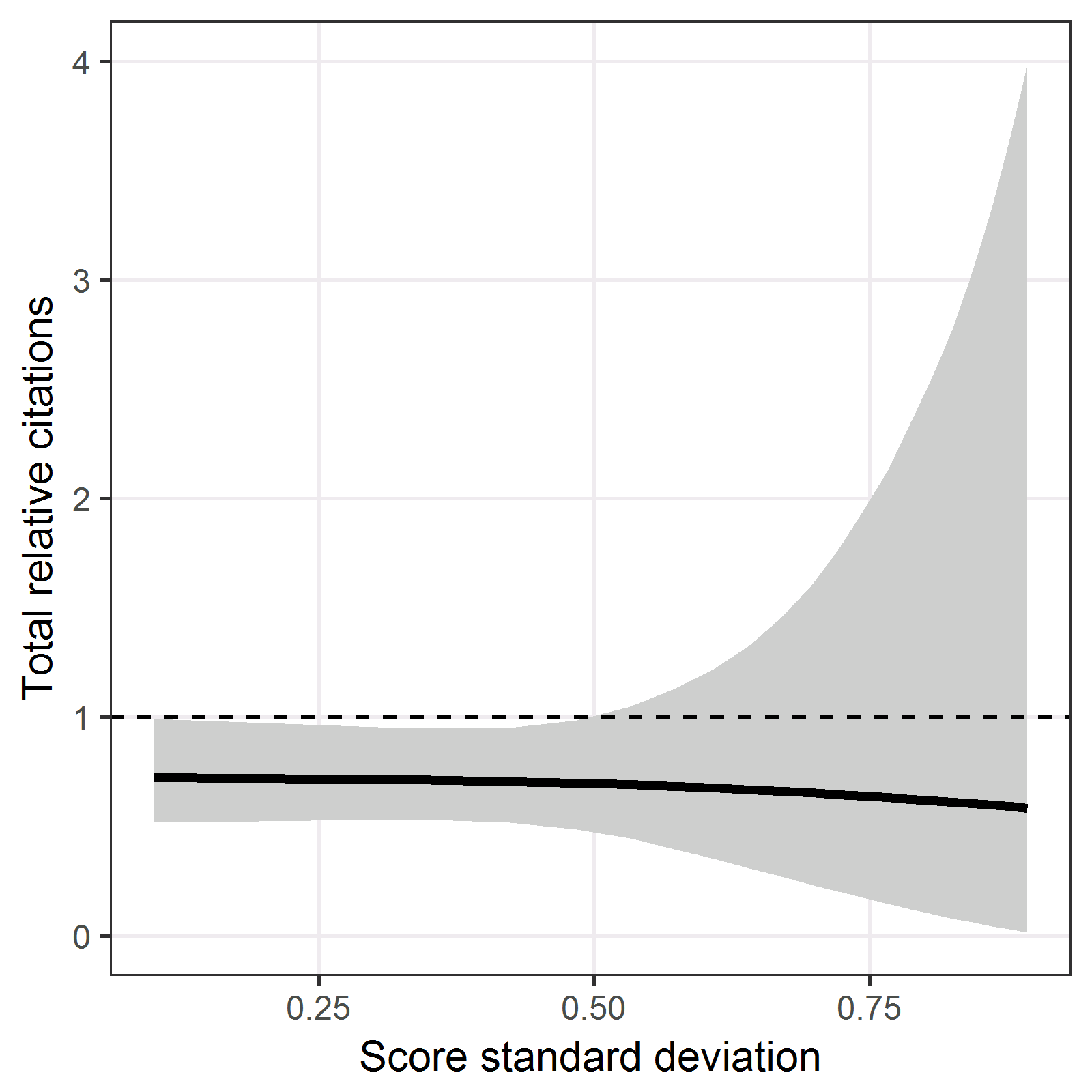
There is one clear outlier using Cook’s distance which may be an influential observation. This is the application with the largest standard deviation which had a citation number just above zero.

#### Results excluding one influential observation

=========================================  
 Dependent variable:   
 --------------------  
 log(`Web TRC` + 0.1)  
-----------------------------------------  
Mean score, f(-0.5) 5.7 (2.5 to 9.0)   
Score SD, f(3) -0.3 (-2.9 to 2.4)   
Review year, f(-0.5) -1.2 (-3.0 to 0.7)   
Intercept -3.9 (-6.8 to -1.1)   
-----------------------------------------  
Observations 226   
Log Likelihood -436.9   
=========================================  
Note:

The estimate for the score standard deviation is very close to zero, indicating little influence of the standard deviation on citations.

##### Predictions for standard deviation



The mean line is flat, showing no clear association between the standard deviation and total citations. There is still uncertainty for high standard deviations, even after excluding the influential observation, although this has been reduced compared with the previous figure.

### Model using score range

Here we use the range in place of the standard deviation as the measure of panel disagreement.

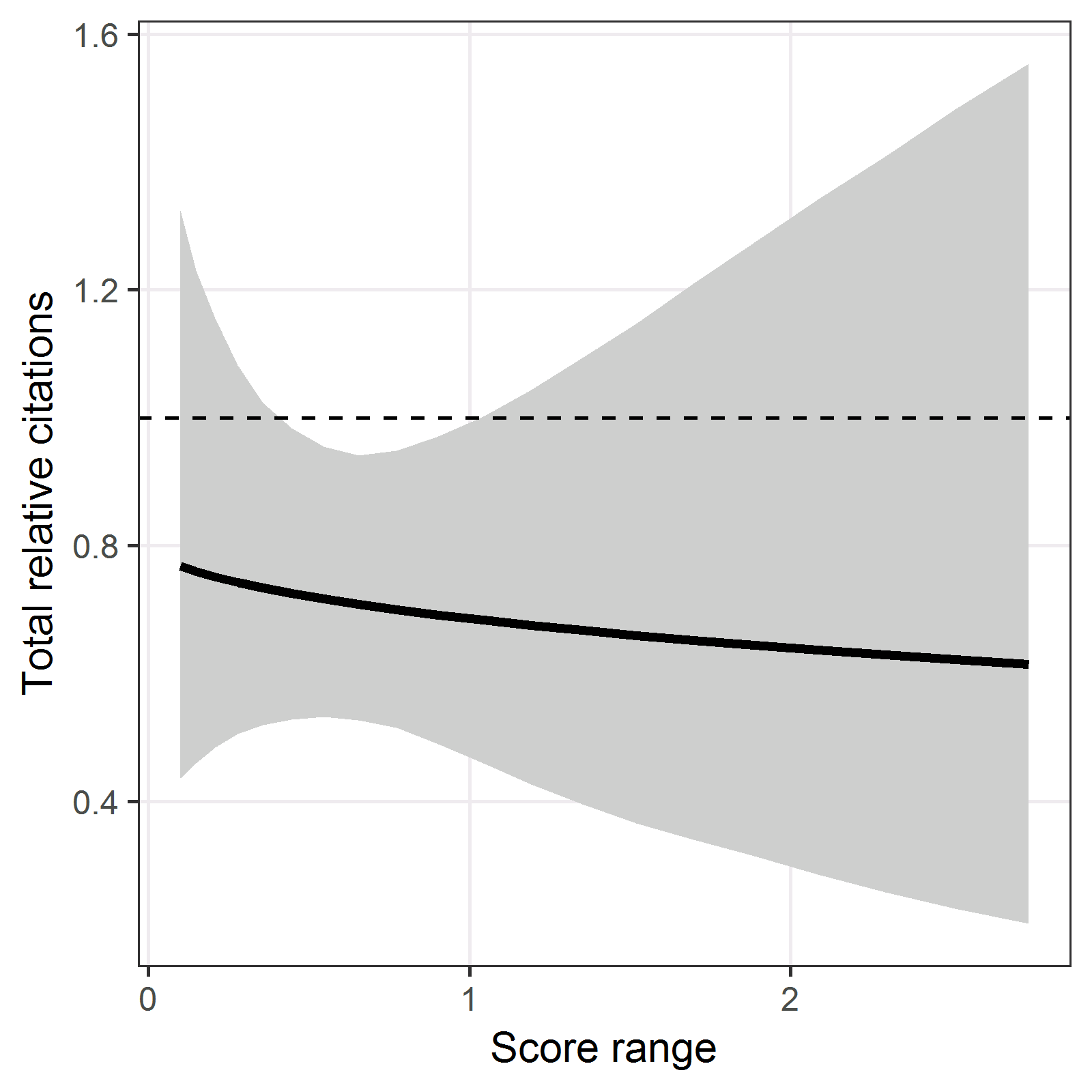
#### Model fit - Top five deviances for score range (FP = Fractional Polynomial)

|  |  |  |  |
| --- | --- | --- | --- |
| FP1 | FP2 | FP3 | Deviance.mean |
| -0.5 | 0.5 | -0.5 | 879.9 |
| -0.5 | 0.5 | -1 | 880.3 |
| -0.5 | 0.5 | -2 | 880.5 |
| -0.5 | -0.5 | -0.5 | 880.6 |
| 0.5 | 0.5 | -0.5 | 880.9 |

#### Table of parameter estimates

=========================================  
 Dependent variable:   
 --------------------  
 log(`Web TRC` + 0.1)  
-----------------------------------------  
Mean score, f(-0.5) 5.4 (2.1 to 8.7)   
Score range, f(0.5) -0.1 (-1.0 to 0.8)   
Review year, f(-0.5) -1.2 (-3.0 to 0.6)   
Intercept -3.6 (-6.6 to -0.5)   
-----------------------------------------  
Observations 227   
Log Likelihood -439.9   
=========================================  
Note:

##### Predictions for range



There was no clear association between the range in an application’s score and its subsequent citation numbers.

## References

* Gallo SA, Carpenter AS, Irwin D, McPartland CD, Travis J, et al. (2014) The Validation of Peer Review through Research Impact Measures and the Implications for Funding Strategies. *PLOS ONE* **9**(9): e106474. <https://doi.org/10.1371/journal.pone.0106474>
* Stef van Buuren, Karin Groothuis-Oudshoorn (2011) mice: Multivariate Imputation by Chained Equations in R. *J Stat Software* **45**(3)