Testosterone, diversity, and group project performance

Cathy Su

9/10/2019

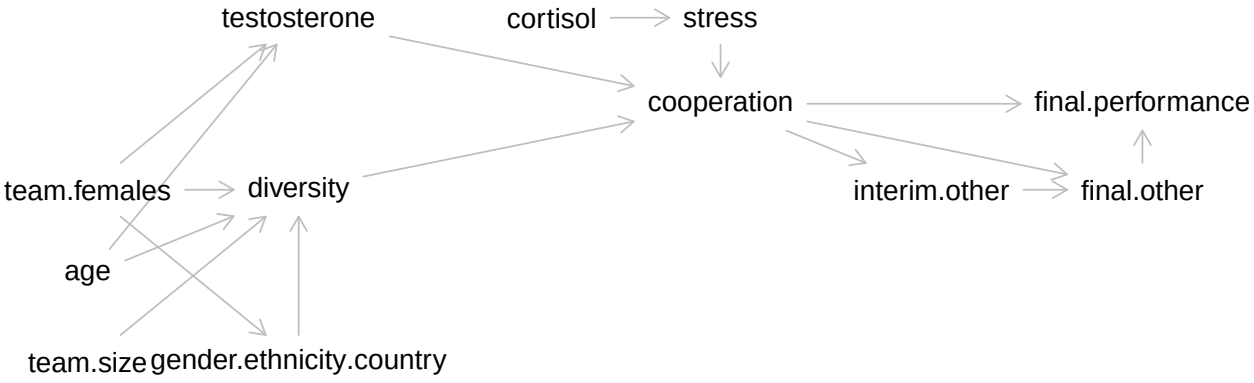
Introduction

We have both individual-level and team-level data for MBA students working on a competitive task (Akinola et al. 2018), processed at Nifty Datasets. There are 370 individuals organized into 74 teams. Our objective in the project is to further examine the relationship between testosterone and group performance, adopting the authors' hypothesis that:

> diversity is beneficial for performance, but only if group-level testosterone is low; diversity has a negative effect on performance if group-level testosterone is high.

Causal diagram

Based on the preamble from (Akinola et al. 2018) we may guess that the effects of testosterone and diversity on performance are mediated by their opposite effects on 'cooperation' (not directly measured) in the group. Furthermore cortisol levels largely unevaluated by the study may influence performance through affecting group 'stress' (not directly measured). Putting this together with the measured variables, our hypothesized causal diagram follows:



Methods

Handling missing data

Before calculating additional team level statistics, we saw there were <10 individuals with missing data (shown below). We preferred to not remove any individuals since we are trying to look at team level performance. For these individuals, not everything was missing so we calculated group average measurements, e.g. average hormone measurements, from other members. When we did this, there were no teams with missing data aside from in the 'interim' variables.

```
##      ID team.id Age Gender   Ethnicity Cortisol Testosterone log.cortisol
## 22  140      16  26 Female       Black       NA           NA           NA
## 45  218       2  NA   Male       Other    0.126       122.44    -2.071473
## 53  236       3  NA   Male       Other    0.086        66.54    -2.453408
## 95  346      36  28   Male       White       NA           NA           NA
## 113 420      47  30   Male South Asian       NA           NA           NA
## 180 546      97  NA   Male       Asian    0.350       132.65    -1.049822
##     log.testosterone Country
## 22                NA     USA
## 45          4.807621
## 53          4.197803
## 95                NA  Canada
## 113               NA   India
## 180         4.887714   Korea
```

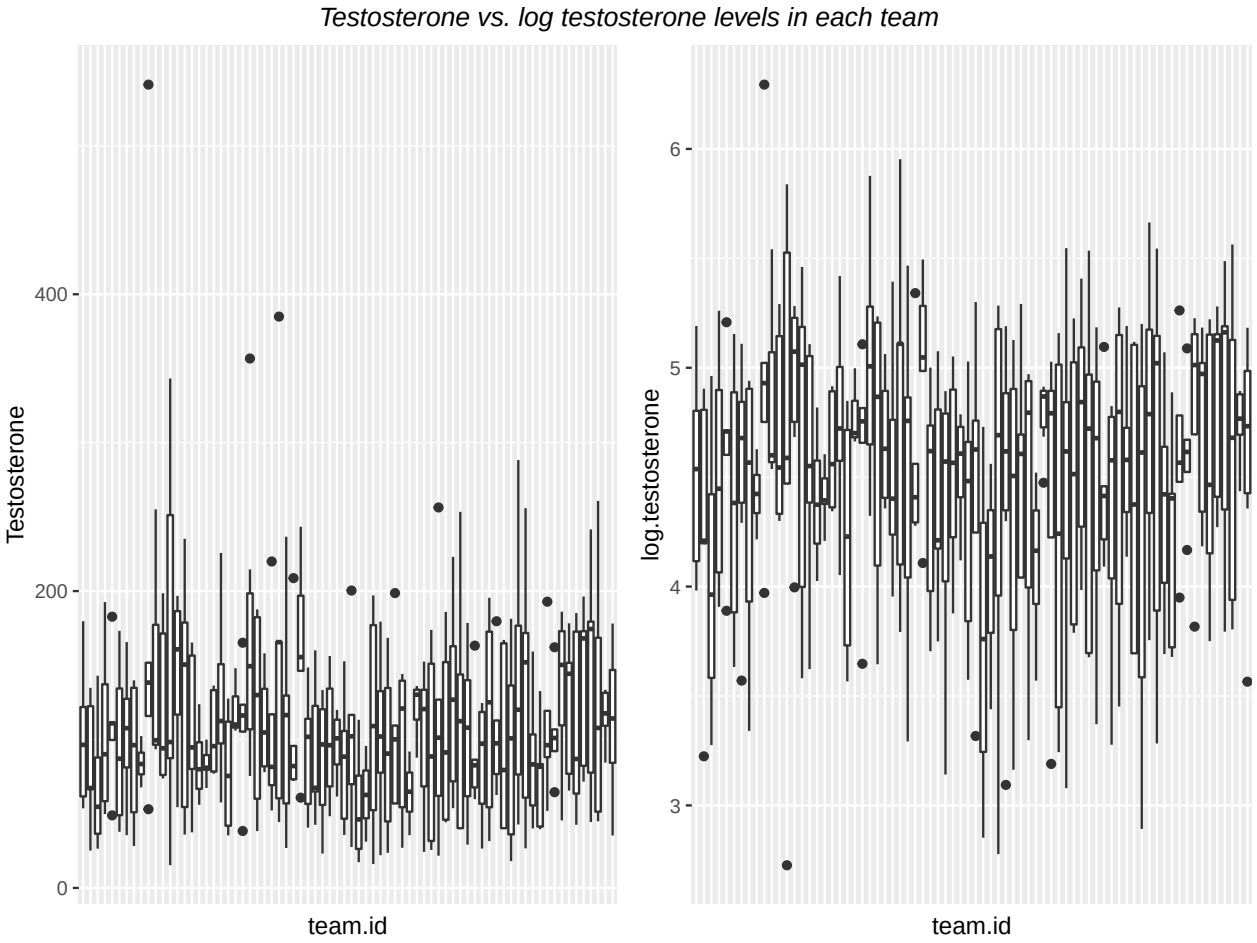
Calculation of group level variables from individual level variables

We are interested in doing our analysis at the group level therefore we needed to calculate the group level values for the diversity score, average log testosterone level, average log cortisol level, average age and average age variance. For the sake of making the units clear we show the code below. With this we can proceed to data exploration.

```
get_score <-function(group){
  score <- length(unique(ind_dat$combo[ind_dat$team.id == group,]))
  return(score)
}
#  calculate the number of unique gender-ethnicity-country combinations
ind_dat$combo <-paste(ind_dat$Gender, ind_dat$Ethnicity, ind_dat$Country)
team_dat$score<- unlist(lapply(team_dat$team.id,
               function(x){length(unique(ind_dat$combo[ind_dat$team.id == x]))}))

#  calculate the average testosterone level for each group.
# some have missing testosterone data, this means we need to average.
team_dat$avg.log.testosterone<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$log.testosterone[(ind_dat$team.id == x)], na.rm = TRUE)}))

#  calculate the average cortisol level for each group.
team_dat$avg.log.cortisol<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$log.cortisol[ind_dat$team.id == x], na.rm = TRUE)}))

#  calculate the average age for each group.
team_dat$avg.age<- unlist(lapply(team_dat$team.id,
               function(x){mean(ind_dat$Age[ind_dat$team.id == x], na.rm = TRUE)}))

team_dat$age.variance<- unlist(lapply(team_dat$team.id,
               function(x){var(ind_dat$Age[ind_dat$team.id == x], na.rm = TRUE)}))
```

Log vs raw values of testosterone and cortisol

Since we are interested in how the the levels of these two hormones may be related to the other variables, we first wanted to check whether we should use their log vs raw values in our EDA.

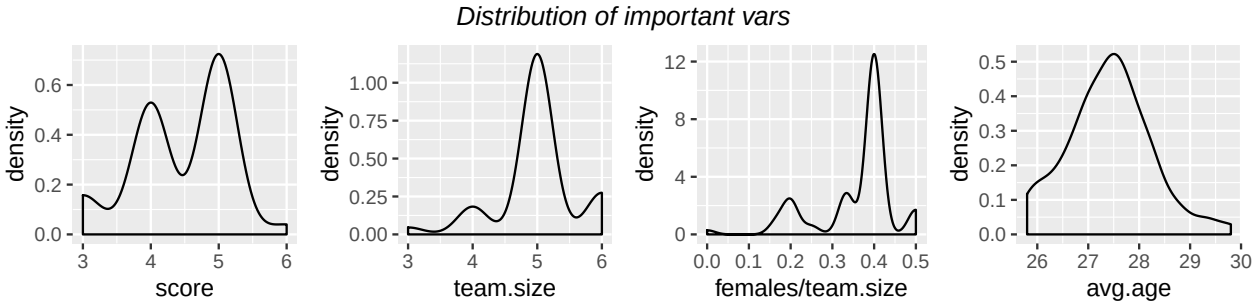


(Cortisol looks similar to the above, except for scale.) It was clear when for both hormones that the log transformed values were distributed with less skew across teams which would be our preference. The authors also chose to use averaged log testosterone per group.

Exploratory Data Analysis & Data Summary

Distributions of variables

We checked the distribution of other key variables besides hormone level in our model, and saw that in particular, our diversity score and number of females per group appear bimodal where team size and age are not. Although our score is differently calculated, this agrees with (Akinola et al. 2018) which classified diversity score into two bins.



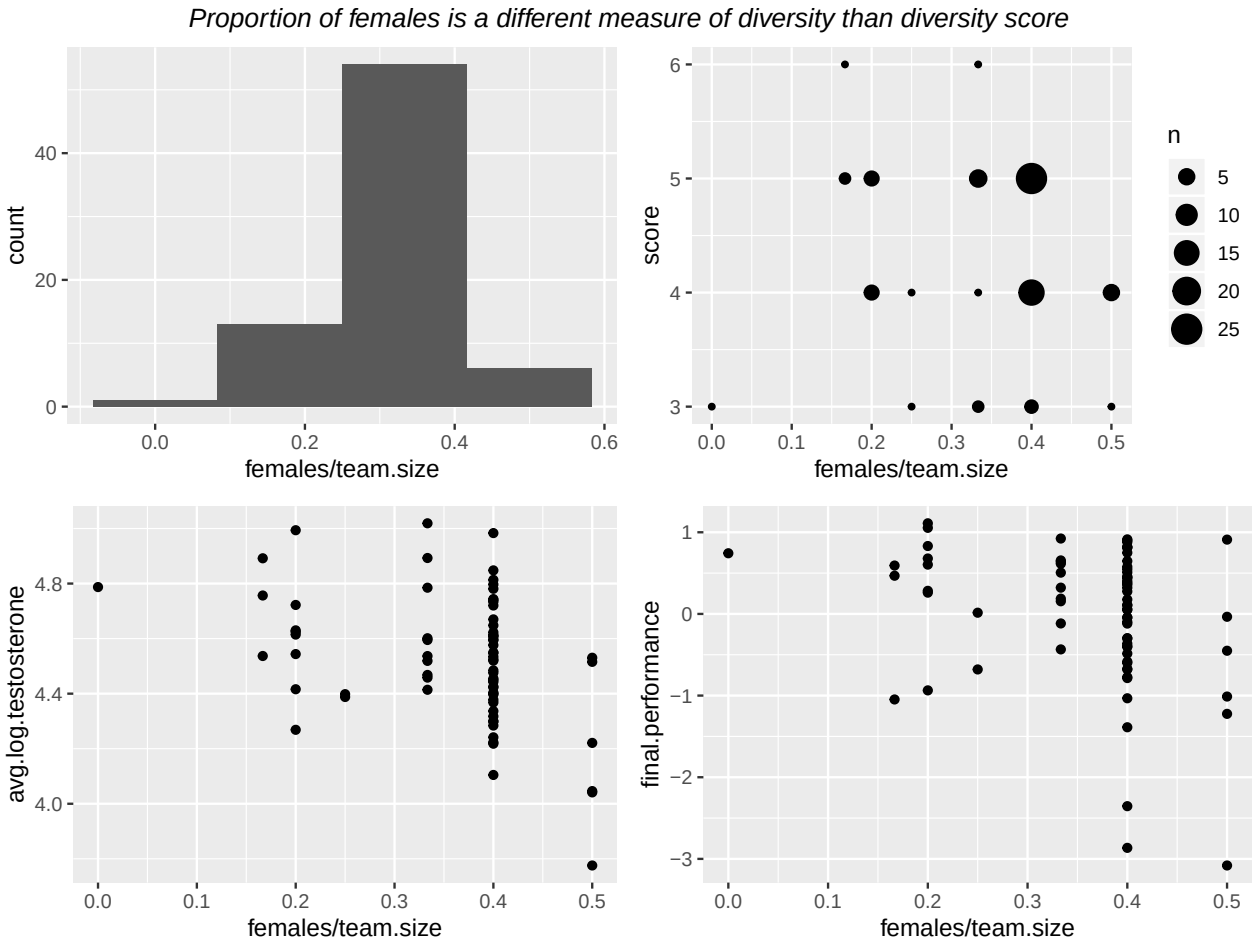
Variable selection

To choose what other variables to include in further analysis, we visually inspected pairwise correlations. We removed the following:

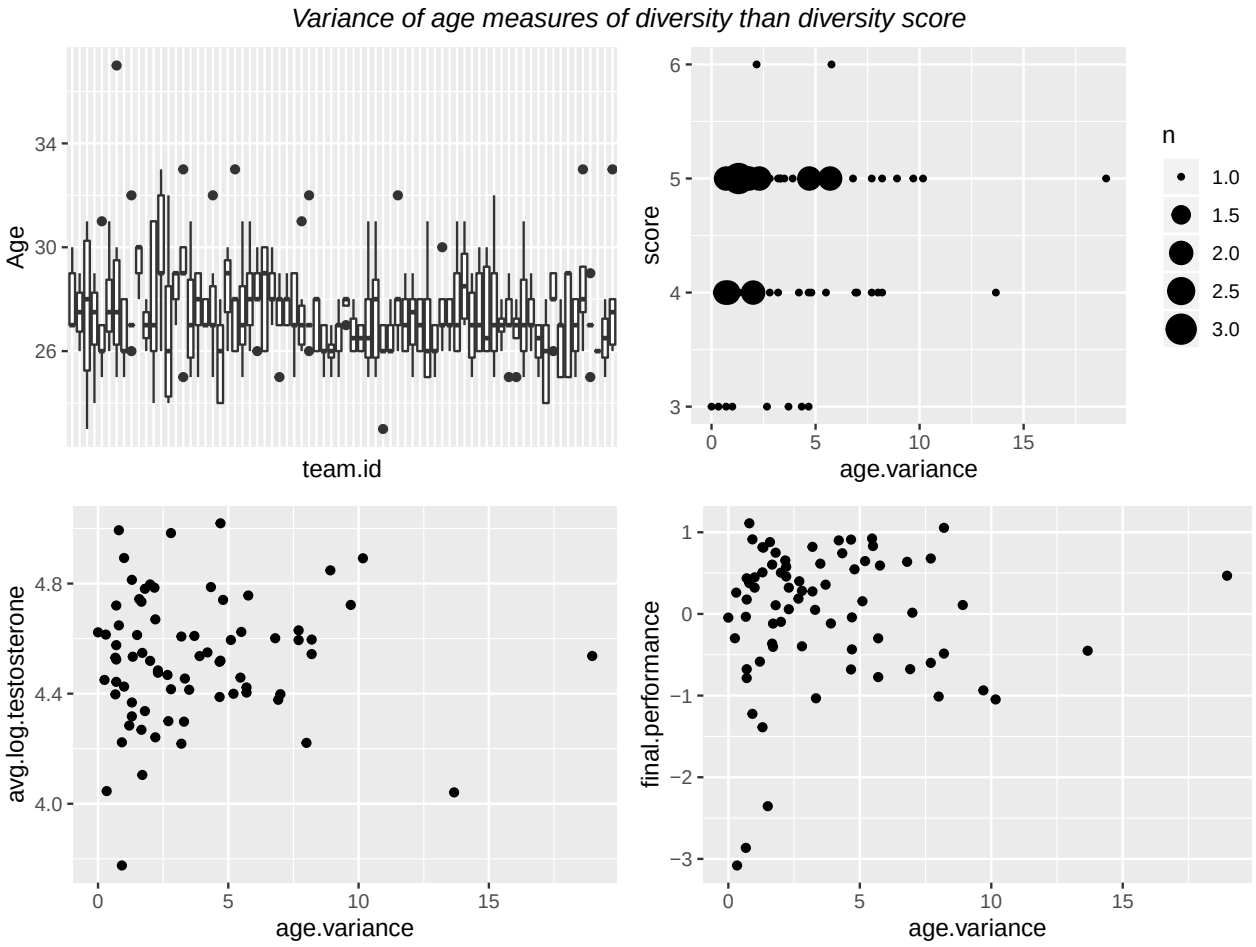
- time.of.day : didn't seem to have a straightforward relationship with other variables.
- 'interim'variables: contain missing data for many teams.
- other 'final' variables besides final.performance: these variables are generally correlated, but not in a straightforward way. The original study states that these measures were standardized and then averaged to form the final.performance score. Since the judges who assigned the final.performance are also domain experts, we have decided to discard the other variables.

We then looked more closely at the variables 'females','age' and 'cortisol'. The former two may contribute to diversity outside of diversity score and the latter may be useful hormone which influences cooperation.

Gender is incorporated in the diversity score, but the variable females was measured separately. Since we are looking at average hormone scores, we examined the proportion of females per group. As seen below, proportion of females is not clearly correlated with diversity score yet shows similar negative correlation with both testosterone and team.size. Our hypothesized causal graph suggested proportion of females should influence performance through both diversity and through affecting testosterone level. However since the two scatterplots on the RHS are so similar, proportion of females seems may not affect performance through diversity score but rather through their effect on testosterone.

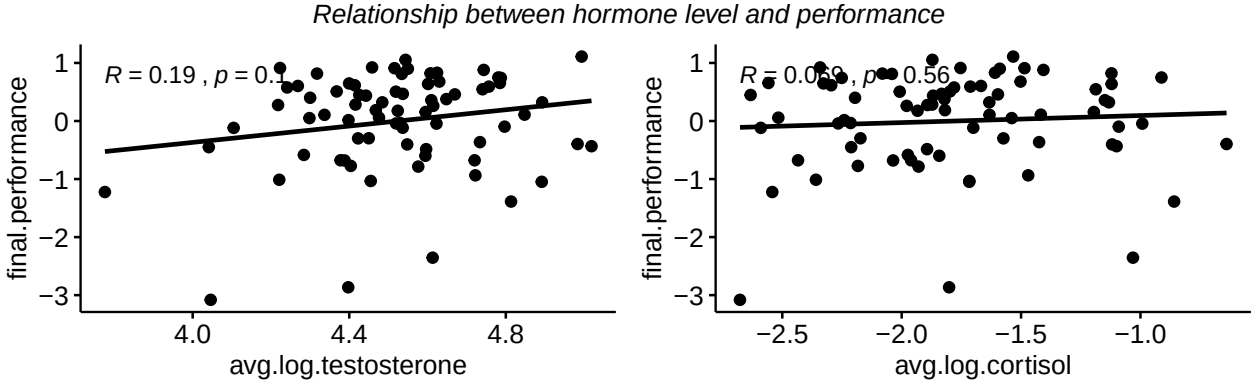


A similar situation occurs for age. When we plotted the spread of age across groups (below left), it seemed mean age is not too different between groups. Since according to our causal graph, the effects of age may influence performance through changing cooperation, we further plotted the variance of age since different ages in a group could also lead to conflict. It also doesn't correlate clearly with diversity score but does show some positive correlation with testosterone and performance. Once again, since the two scatterplots on the RHS are so similar, variance and mean of age may not affect performance through diversity score but rather through their effect on testosterone.



Based on the above, we decided to keep discard 'age' and 'females' as extra variables which may tell us which may not necessarily tell us more about the relationship between testosterone and performance. However, we can later choose to use them if diversity score alone is insufficient.

Lastly, the correlation of cortisol with final.performance seems weaker than the correlation of testosterone with final.performance. However this



Results

The results presented by the original study (Akinola et al. 2018) include that:

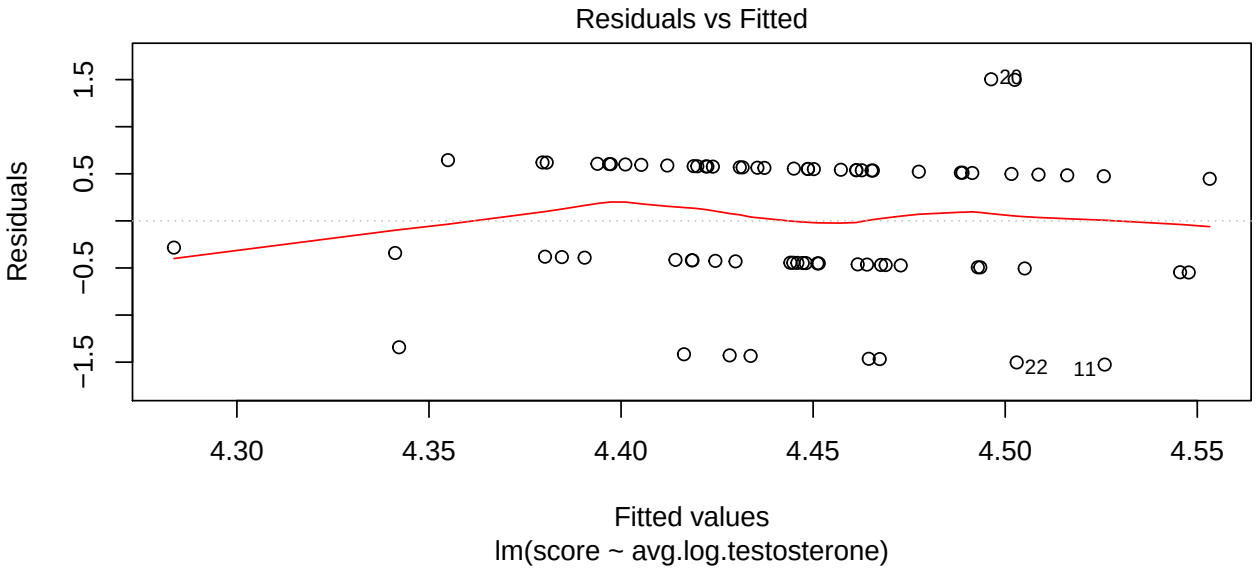
- the interaction between group testosterone and group diversity was positive and significant at $p < .01$

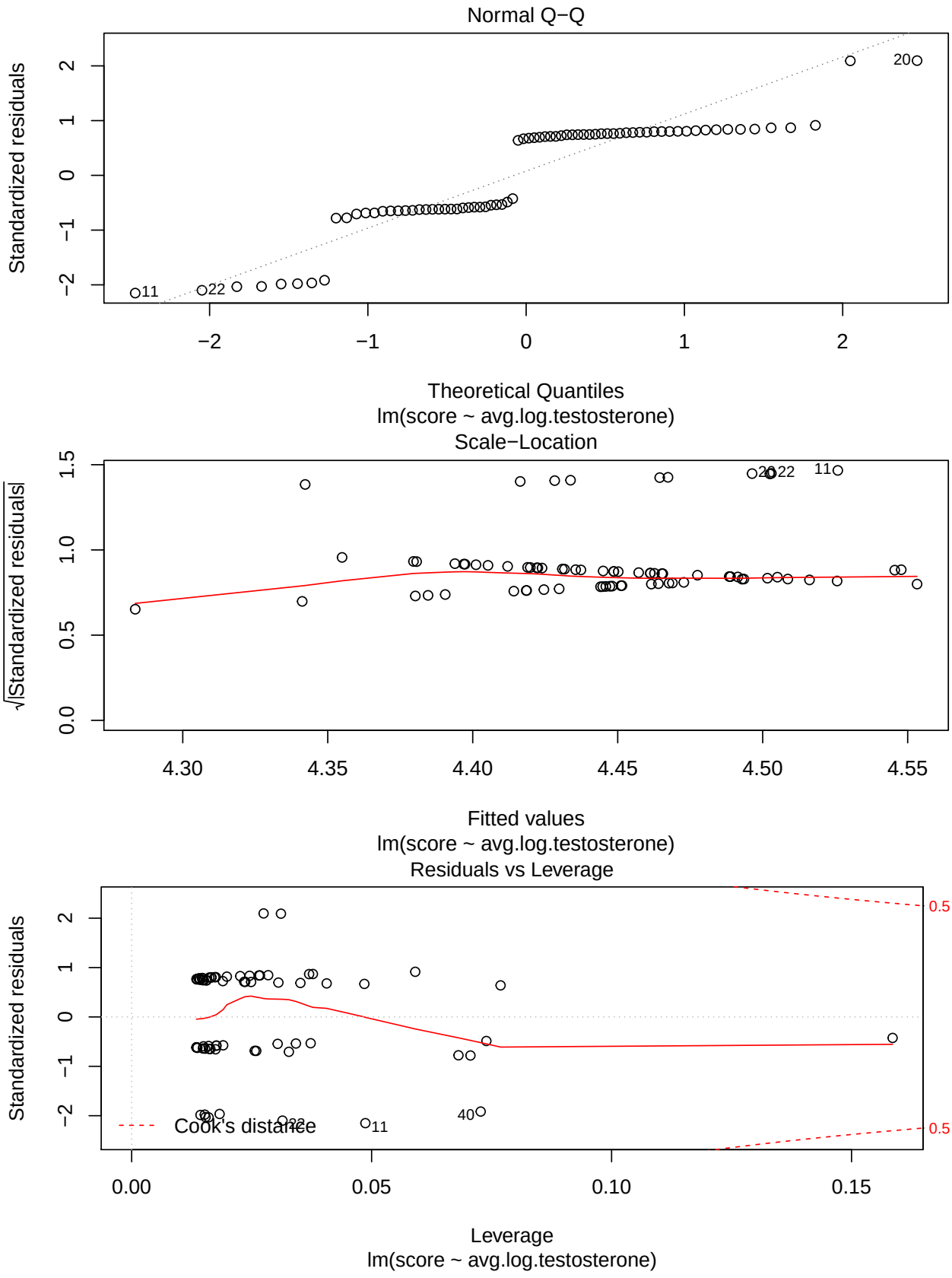
- when group diversity was low, group testosterone significantly positively predicted performance at $p < .01$,
- when group diversity was relatively high (Fau score was 1 SD below the mean), group testosterone significantly negatively predicted performance $p < .01$,

interaction between group testosterone and group diversity

To test the author's hypothesis, we checked whether we could find the interaction between group testosterone and group diversity. However it seems like our non-binary measurement of the group diversity score is not significantly predictive of testosterone.

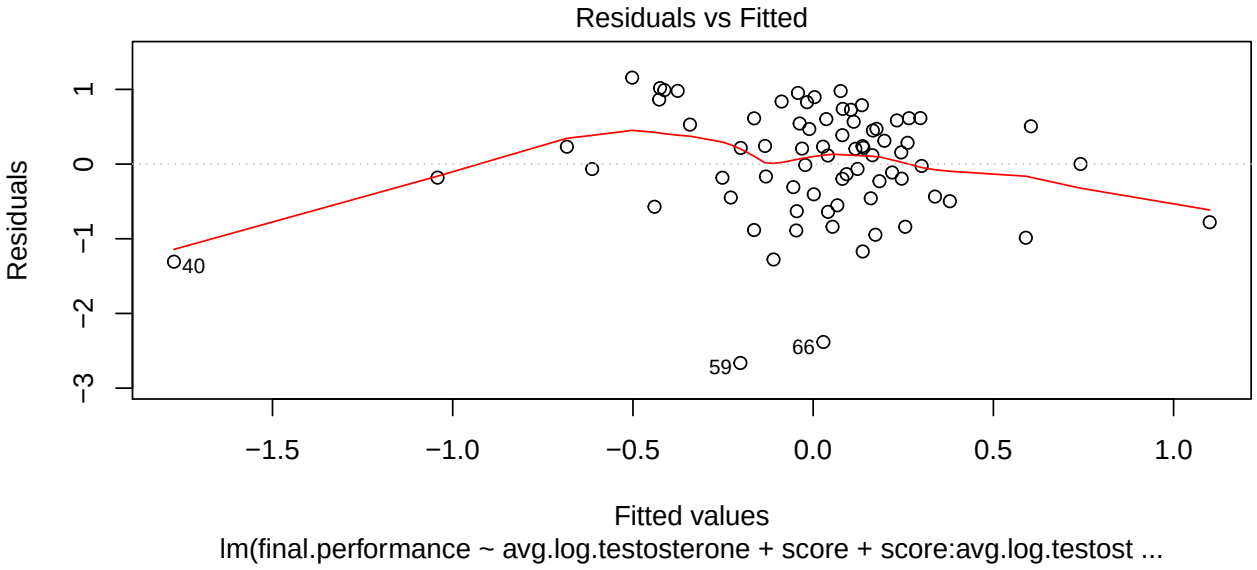
```
##
## Call:
## lm(formula = score ~ avg.log.testosterone, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5259 -0.4514  0.4791  0.5641  1.5036
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.4647     1.6767   2.066   0.0424 *
## avg.log.testosterone   0.2169     0.3701   0.586   0.5597
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7275 on 72 degrees of freedom
## Multiple R-squared:  0.004746,   Adjusted R-squared:  -0.009077
## F-statistic: 0.3434 on 1 and 72 DF,  p-value: 0.5597
```

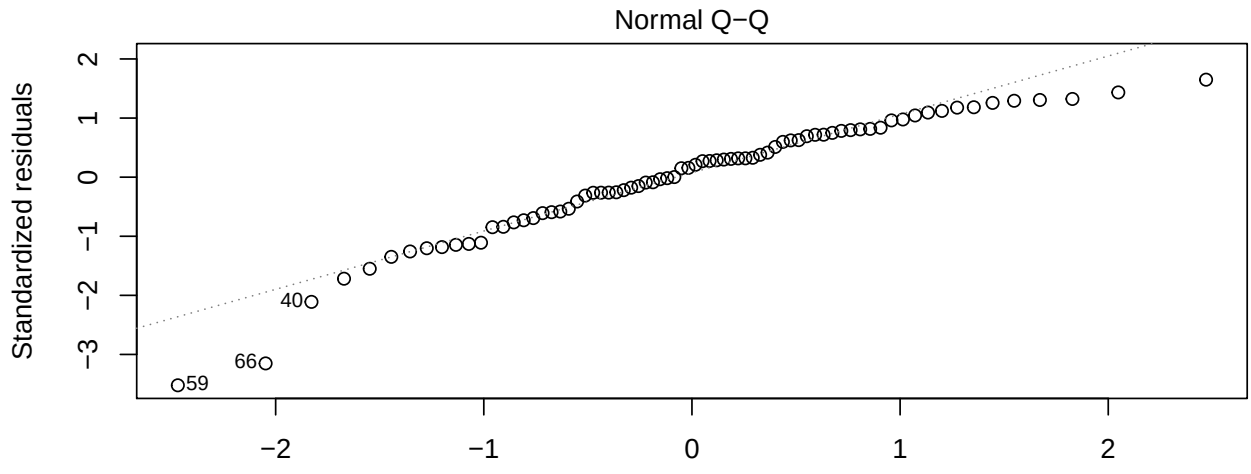


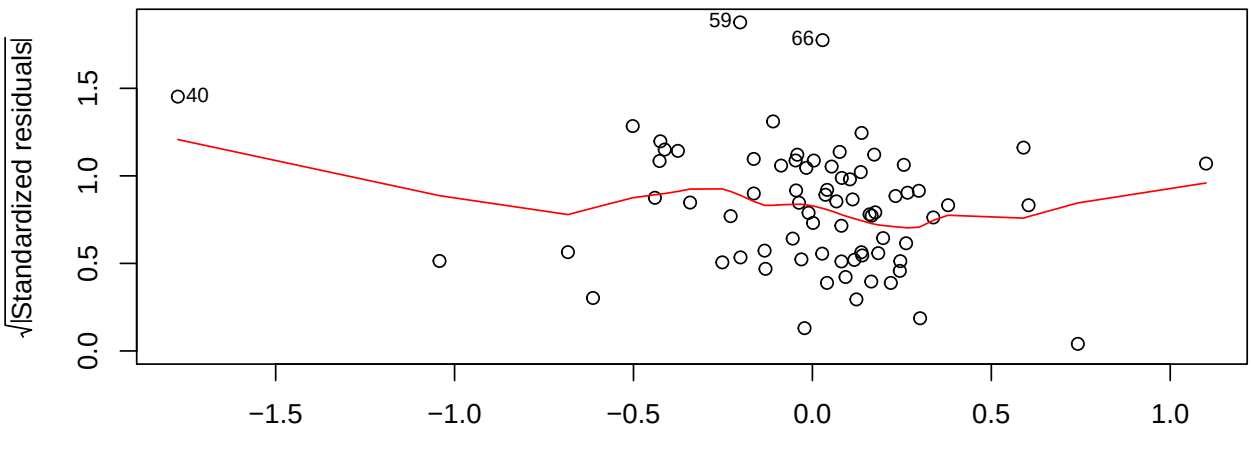

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(score ~ avg.log.testosterone)
Scale−Location
√|Standardized residuals|
Fitted values
lm(score ~ avg.log.testosterone)
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(score ~ avg.log.testosterone)

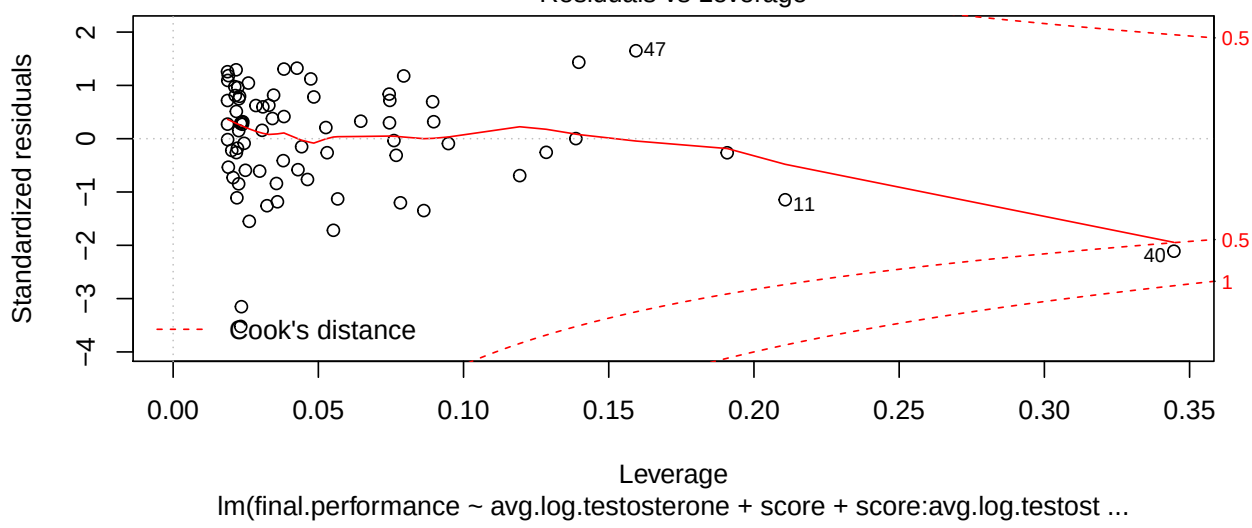
effect of group diversity on relationship between testosterone and performance

```
##
## Call:
## lm(formula = final.performance ~ avg.log.testosterone + score +
##     score:avg.log.testosterone, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6630 -0.4447  0.1371  0.5607  1.1573
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -43.5696    10.8017  -4.034 0.000138 ***
## avg.log.testosterone         9.5172     2.3692   4.017 0.000146 ***
## score                        9.3562     2.4672   3.792 0.000314 ***
## avg.log.testosterone:score  -2.0415     0.5406  -3.777 0.000330 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7651 on 70 degrees of freedom
## Multiple R-squared:  0.2017, Adjusted R-squared:  0.1675
## F-statistic: 5.897 on 3 and 70 DF,  p-value: 0.001198
```



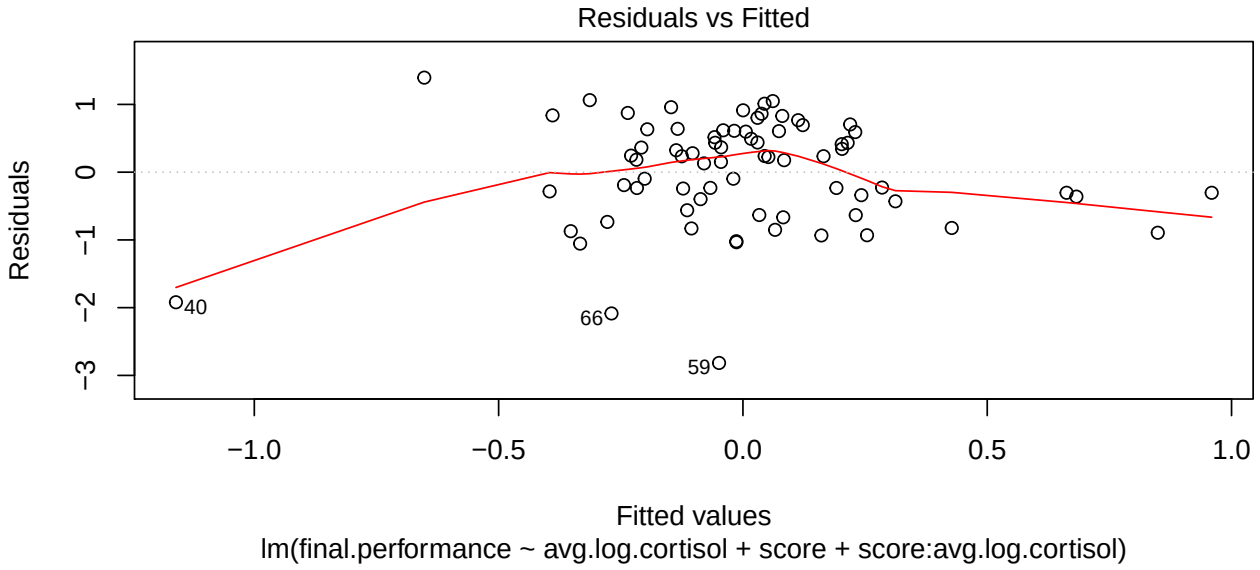

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(final.performance ~ avg.log.testosterone + score + score:avg.log.testost ...
40
66
59

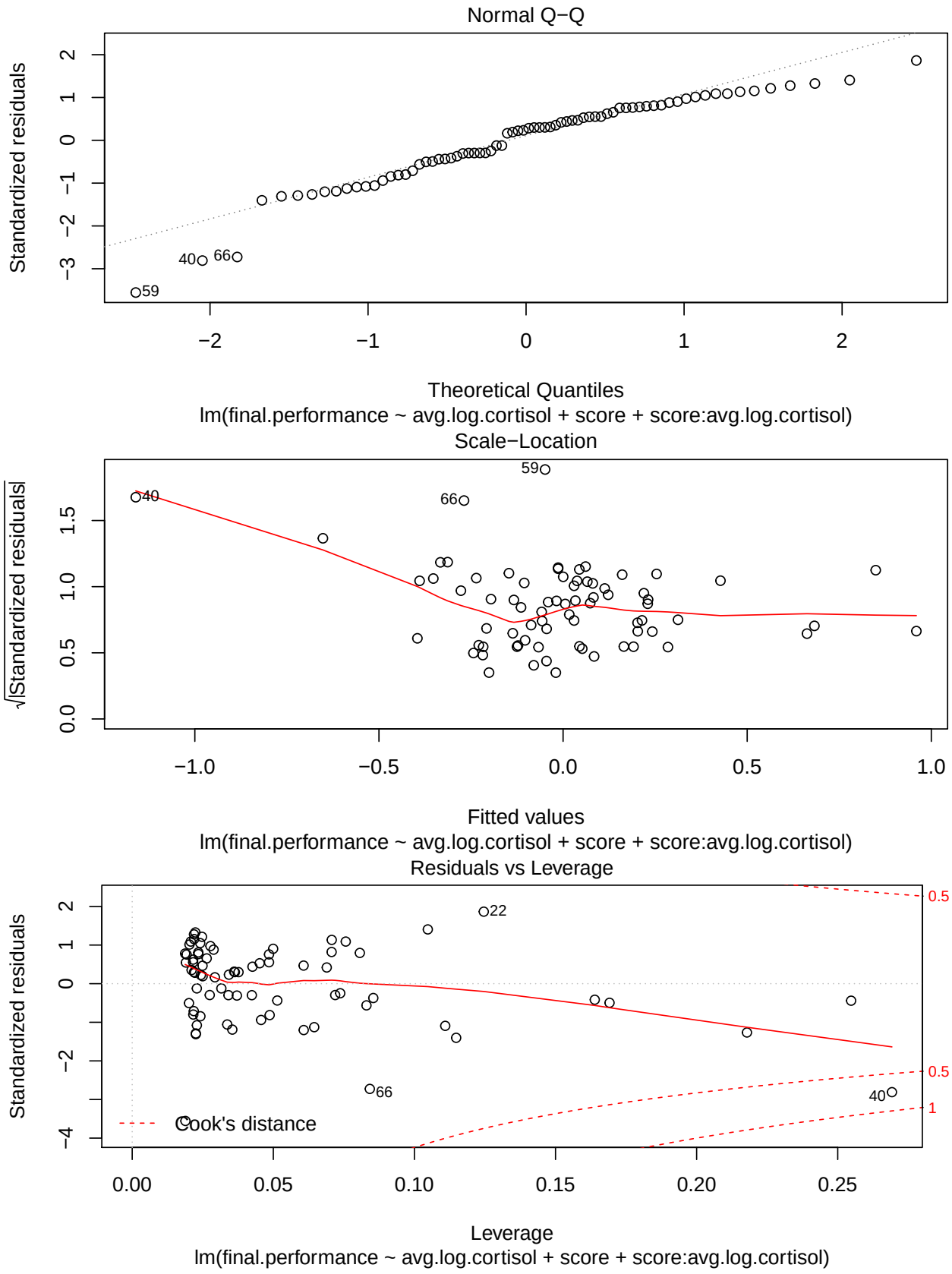

Scale−Location
√|Standardized residuals|
Fitted values
lm(final.performance ~ avg.log.testosterone + score + score:avg.log.testost ...
59
66
40


Residuals vs Leverage
Standardized residuals
Leverage
lm(final.performance ~ avg.log.testosterone + score + score:avg.log.testost ...
Cook's distance
0.5
0.5
1
47
11
40

effect of cortisol on relationship between diversity and performance

```
##
## Call:
## lm(formula = final.performance ~ avg.log.cortisol + score + score:avg.log.cortisol,
##     data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8160 -0.4223  0.2032  0.6035  1.3951
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               6.0611     2.0908   2.899  0.00500 **
## avg.log.cortisol          3.5407     1.1183   3.166  0.00229 **
## score                    -1.3429     0.4742  -2.832  0.00604 **
## avg.log.cortisol:score   -0.7827     0.2525  -3.100  0.00279 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7996 on 70 degrees of freedom
## Multiple R-squared:  0.1282, Adjusted R-squared:  0.09082
## F-statistic: 3.431 on 3 and 70 DF,  p-value: 0.02159
```

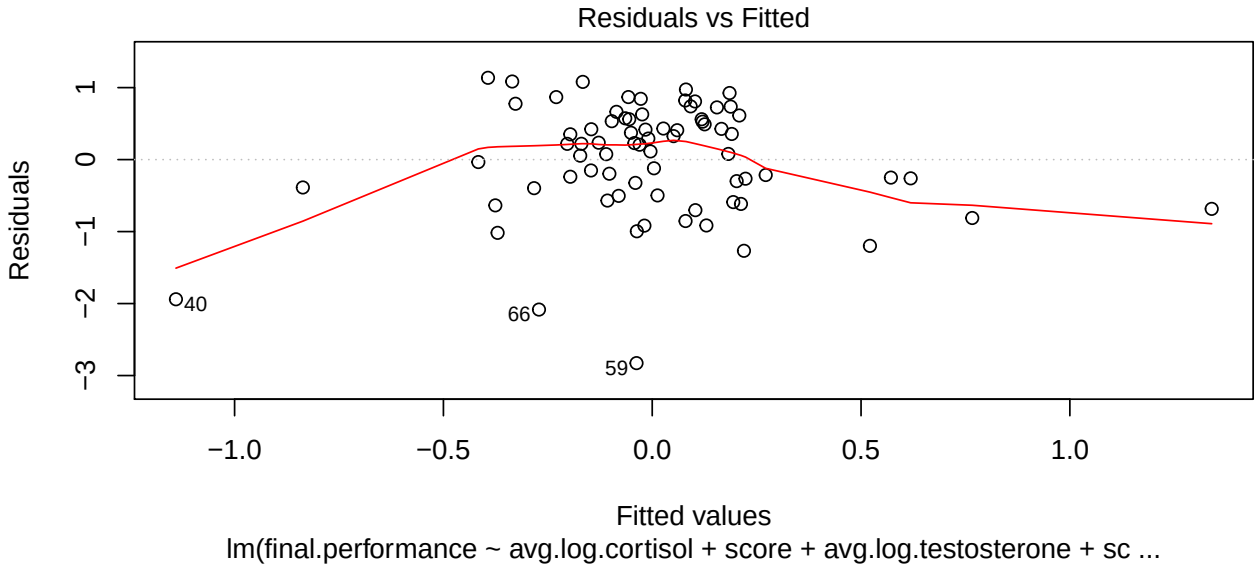


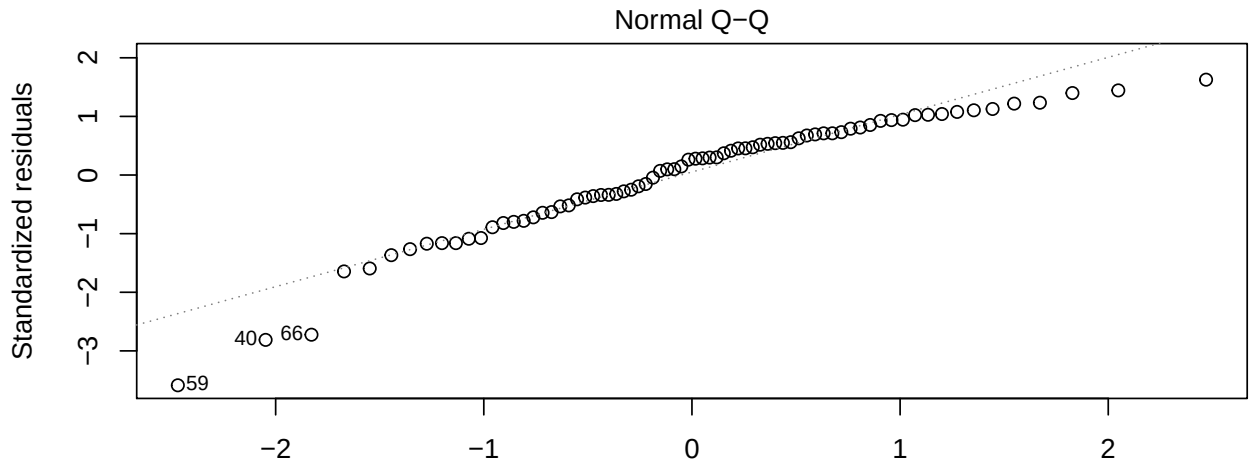


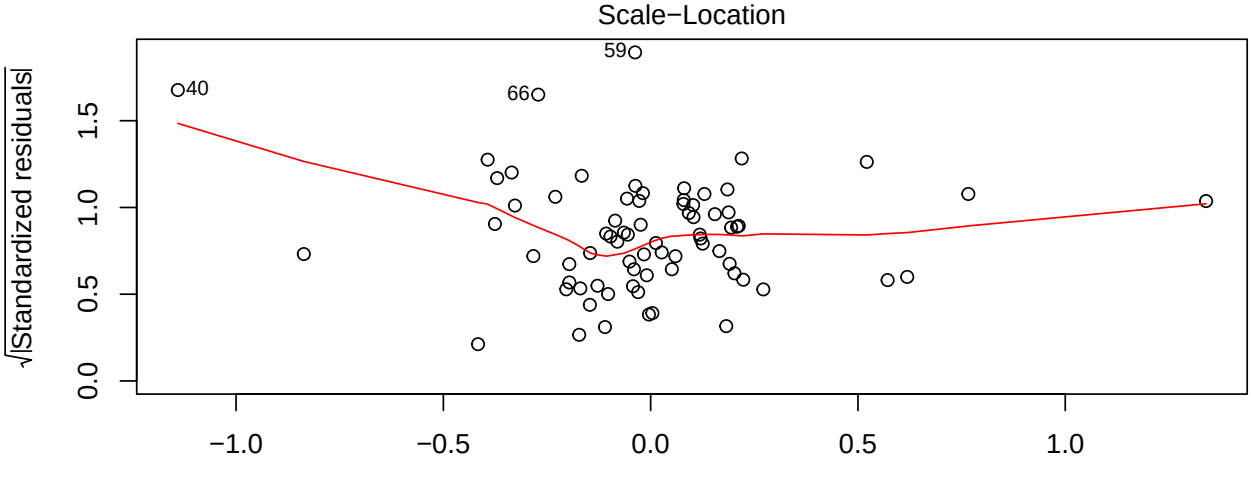
Surprisingly, when substituting cortisol for testosterone, it seems that

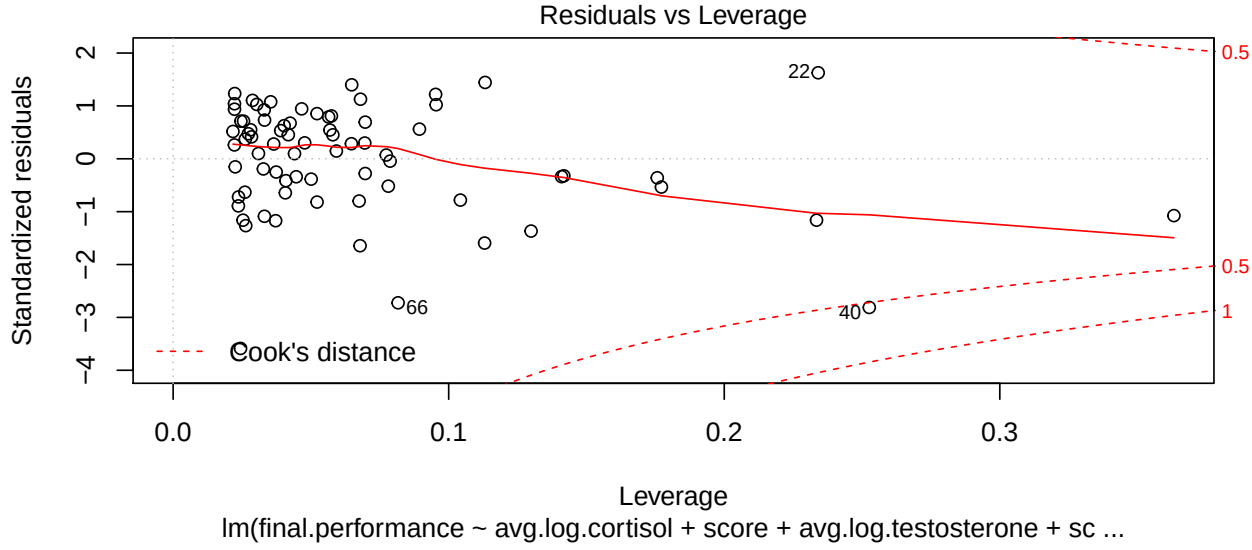
Combination of cortisol, performance, and

```
##
## Call:
## lm(formula = final.performance ~ avg.log.cortisol + score + avg.log.testosterone +
##     score:avg.log.cortisol:avg.log.testosterone, data = team_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8273 -0.4727  0.2126  0.5606  1.1362
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                                 8.9460     5.2055   1.719
## avg.log.cortisol                            3.1283     1.1520   2.716
## score                                      -1.2450     0.4669  -2.667
## avg.log.testosterone                       -0.7554     0.7453  -1.014
## avg.log.cortisol:score:avg.log.testosterone -0.1563     0.0544  -2.873
##                                           Pr(>|t|)
## (Intercept)                                0.09018 .
## avg.log.cortisol                           0.00835 **
## score                                      0.00954 **
## avg.log.testosterone                       0.31434
## avg.log.cortisol:score:avg.log.testosterone  0.00539 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.798 on 69 degrees of freedom
## Multiple R-squared:  0.1439, Adjusted R-squared:  0.09426
## F-statistic: 2.899 on 4 and 69 DF,  p-value: 0.02805
```




Normal Q−Q
Standardized residuals
40
66
59
Theoretical Quantiles
lm(final.performance ~ avg.log.cortisol + score + avg.log.testosterone + sc ...


Scale−Location
√|Standardized residuals|
59
40
66
Fitted values
lm(final.performance ~ avg.log.cortisol + score + avg.log.testosterone + sc ...


Residuals vs Leverage
Standardized residuals
22
66
40
Cook's distance
0.5
0.5
1
Leverage
lm(final.performance ~ avg.log.cortisol + score + avg.log.testosterone + sc ...

Conclusion

TBD

Bibliography

Akinola, Modupe, Elizabeth Page-Gould, Pranjal H. Mehta, and Zaijia Liu. 2018. "Hormone-Diversity Fit: Collective Testosterone Moderates the Effect of Diversity on Group Performance." *Psychological Science* 29 (6):859–67. https://doi.org/10.1177/0956797617744282.