

0.2 Model Fitting

Anni Hong

9/17/2020

```
combo_dat <- read.csv("./data/combined_processed.csv")
summary(combo_dat)
```

```
##      team.id      team.size final.performance  time.of.day
## Min.   : 2.00    Min.   :3    Min.   : -3.0807    Min.   : 9.000
## 1st Qu.: 34.25   1st Qu.:5    1st Qu.: -0.4267   1st Qu.: 9.438
## Median : 62.50   Median :5    Median : 0.1817   Median :10.750
## Mean   : 60.08   Mean   :5    Mean   : 0.0000   Mean   :11.672
## 3rd Qu.: 86.75   3rd Qu.:5    3rd Qu.: 0.6012   3rd Qu.:14.250
## Max.   :111.00   Max.   :6    Max.   : 1.1099   Max.   :16.000
##
##      females      final.cash      final.contracts final.reorders
## Min.   :0.000    Min.   : 642783    Min.   :1.000    Min.   : 15.00
## 1st Qu.:2.000    1st Qu.:1362974   1st Qu.:2.000    1st Qu.: 81.25
## Median :2.000    Median :1664432   Median :3.000    Median : 86.00
## Mean   :1.784    Mean   :1600262   Mean   :2.662    Mean   : 84.54
## 3rd Qu.:2.000    3rd Qu.:1820144   3rd Qu.:3.000    3rd Qu.: 90.00
## Max.   :2.000    Max.   :2050636   Max.   :3.000    Max.   :110.00
##
##      final.rank      interim.performance  interim.cash      interim.contracts
## Min.   : 1.000    Min.   : -2.1978    Min.   : 396109    Min.   :1.000
## 1st Qu.: 4.000    1st Qu.: -0.2651    1st Qu.: 734886    1st Qu.:2.000
## Median : 7.500    Median : 0.1456    Median : 806530    Median :3.000
## Mean   : 7.257    Mean   : 0.0000    Mean   : 812429    Mean   :2.404
## 3rd Qu.:10.000    3rd Qu.: 0.6604    3rd Qu.: 925021    3rd Qu.:3.000
## Max.   :14.000    Max.   : 1.0924    Max.   :1062138    Max.   :3.000
##      NA's :22      NA's :22      NA's :22      NA's :22
##      interim.reorders  interim.rank      mean_testo      mean_log_testo
## Min.   : 20.00    Min.   : 1.00    Min.   : 55.75    Min.   :3.776
## 1st Qu.: 75.75    1st Qu.: 4.00    1st Qu.: 95.77    1st Qu.:4.404
## Median : 85.00    Median : 8.00    Median :106.36    Median :4.537
## Mean   : 81.40    Mean   : 8.00    Mean   :111.55    Mean   :4.541
## 3rd Qu.: 90.00    3rd Qu.:11.25    3rd Qu.:122.78    3rd Qu.:4.648
## Max.   :108.00    Max.   :15.00    Max.   :200.04    Max.   :5.019
##      NA's :22      NA's :22      NA's :5      NA's :5
##      sd_testo      sd_log_testo      mean_cortisol      mean_log_cortisol
## Min.   : 12.16    Min.   :0.1471    Min.   :0.0760    Min.   : -2.6326
## 1st Qu.: 45.82    1st Qu.:0.4819    1st Qu.:0.1420    1st Qu.: -2.0813
## Median : 56.86    Median :0.5978    Median :0.1880    Median : -1.8013
## Mean   : 60.98    Mean   :0.6137    Mean   :0.2219    Mean   : -1.7542
## 3rd Qu.: 68.61    3rd Qu.:0.7378    3rd Qu.:0.2540    3rd Qu.: -1.4701
## Max.   :194.45    Max.   :1.2159    Max.   :0.7520    Max.   : -0.6394
```

```
## NA's :5      NA's :5      NA's :5      NA's :5
## sd_cortisol sd_log_cortisol diversity_score team_size
## Min. :0.007762 Min. :0.0699 Min. :3.000 Min. :3
## 1st Qu.:0.058526 1st Qu.:0.3705 1st Qu.:4.000 1st Qu.:5
## Median :0.087864 Median :0.5221 Median :5.000 Median :5
## Mean :0.121867 Mean :0.5143 Mean :4.446 Mean :5
## 3rd Qu.:0.139825 3rd Qu.:0.6486 3rd Qu.:5.000 3rd Qu.:5
## Max. :0.813247 Max. :1.0058 Max. :6.000 Max. :6
## NA's :5      NA's :5
## mean_age proportion_female
## Min. :25.80 Min. :0.0000
## 1st Qu.:27.00 1st Qu.:0.3333
## Median :27.50 Median :0.4000
## Mean :27.47 Mean :0.3586
## 3rd Qu.:27.96 3rd Qu.:0.4000
## Max. :29.80 Max. :0.5000
## NA's :12
```

```
dim(combo_dat)
```

```
## [1] 74 26
```

```
combo_dat <- combo_dat %>%
```

```
  select(final.performance, mean_log_testo, diversity_score, proportion_female, mean_age, team.size, time.of.day)
  drop_na()
```

```
summary(combo_dat)
```

```
## final.performance mean_log_testo diversity_score proportion_female
## Min. : -2.35437 Min. :3.776 Min. :3.000 Min. :0.0000
## 1st Qu.: -0.43444 1st Qu.:4.414 1st Qu.:4.000 1st Qu.:0.3333
## Median : 0.26076 Median :4.544 Median :5.000 Median :0.4000
## Mean : 0.07407 Mean :4.547 Mean :4.474 Mean :0.3491
## 3rd Qu.: 0.61526 3rd Qu.:4.648 3rd Qu.:5.000 3rd Qu.:0.4000
## Max. : 1.10989 Max. :5.019 Max. :6.000 Max. :0.5000
## mean_age team.size time.of.day mean_log_cortisol
## Min. :25.80 Min. :3 Min. :9.00 Min. : -2.6326
## 1st Qu.:27.00 1st Qu.:5 1st Qu.:9.00 1st Qu.: -2.1953
## Median :27.50 Median :5 Median :10.75 Median : -1.8206
## Mean :27.51 Mean :5 Mean :11.67 Mean : -1.7663
## 3rd Qu.:28.00 3rd Qu.:5 3rd Qu.:14.25 3rd Qu.: -1.4856
## Max. :29.80 Max. :6 Max. :16.00 Max. : -0.6394
```

```
dim(combo_dat)
```

```
## [1] 57 8
```

Causal DAG

TODO: - explain the DAG - introduce the variables I added and the variables of interests - write out the stat model - do more testing with the residuals - do more analysis on cortisol - interpreting the results

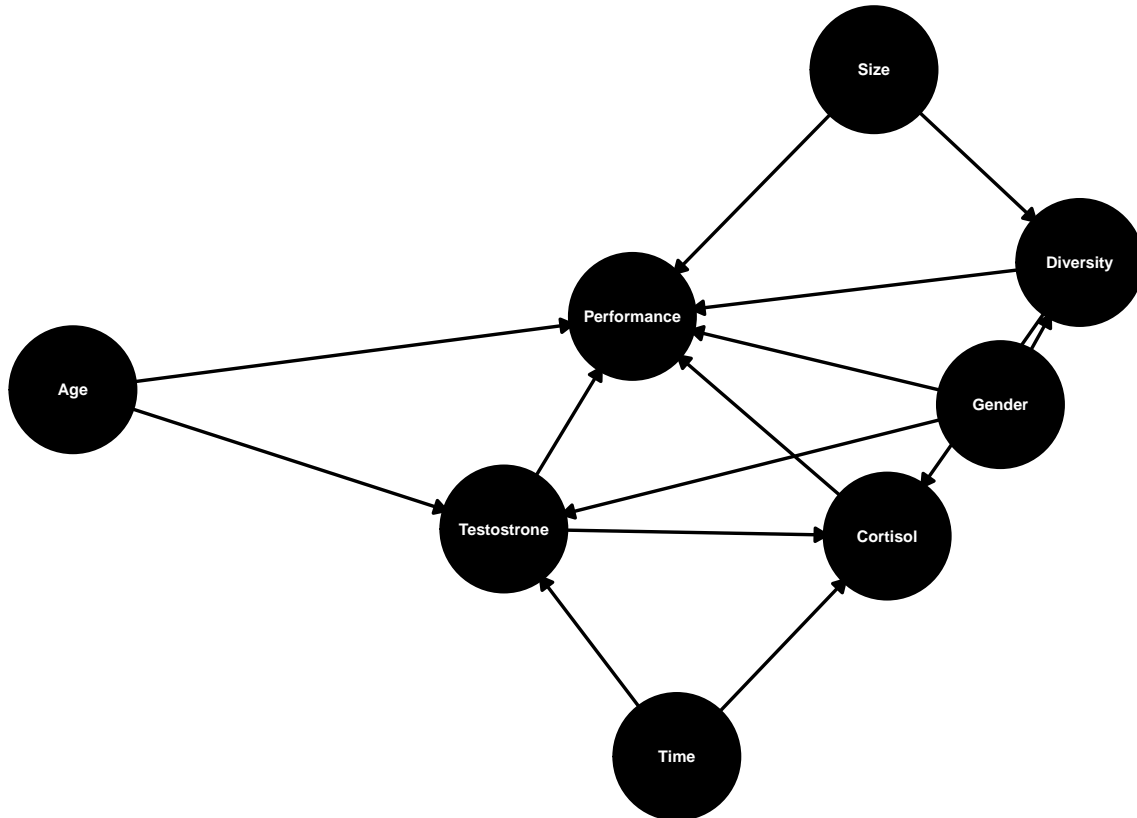
```
tidy_ggdag <- dagify(
  Testosterone ~ Time + Gender + Age,
```

```

Cortisol ~ Time + Testosterone + Diversity,
Diversity ~ Size + Gender,
Performance ~ Testosterone + Cortisol + Diversity + Size + Age + Gender,
exposure = "Diversity",
outcome = "Performance"
) %>%
  tidy_dagitty()

ggdag(tidy_ggdag, node_size = 22, text_size = 2.2) +
  theme_dag()

```



We are interested in the interaction effect of testosterone and diversity on final performance. The original researchers hypothesized that for the group with high testosterone would do worse with higher level of diversity whereas groups with lower testosterone would do better with higher levels of diversity. Therefore base off the DAG drawn above, we would need to condition on gender(calculated as the fraction of female in the group), age (average age of of the group) , Team size, and time of the day that the hormones are sampled. I also hypothesize that testosterone and diversity both have impacts on the cortisol level, maybe more conflicts will arise with higher level of testosterone and diversity, and stress (measured by cortisol) will have an impact on final performance. Thus I intend to test both models, with and without cortisol.

```

model_1 <- lm(data = combo_dat, final.performance ~ mean_log_testo + diversity_score + mean_log_testo*d
summary(model_1)

```

```

##
## Call:
## lm(formula = final.performance ~ mean_log_testo + diversity_score +
##     mean_log_testo * diversity_score + proportion_female + mean_age +
##     team.size + time.of.day, data = combo_dat)

```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9483 -0.3671 -0.0357  0.4744  1.1744
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -35.38923    13.44188   -2.633  0.01130 *
## mean_log_testo     7.95943     2.77072    2.873  0.00600 **
## diversity_score    7.91030     2.86492    2.761  0.00808 **
## proportion_female  -0.78671     1.27714   -0.616  0.54075
## mean_age         -0.09442     0.12059   -0.783  0.43740
## team.size         0.47066     0.18042    2.609  0.01202 *
## time.of.day       0.07790     0.04926    1.582  0.12019
## mean_log_testo:diversity_score -1.79410     0.62822   -2.856  0.00628 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6541 on 49 degrees of freedom
## Multiple R-squared:  0.2588, Adjusted R-squared:  0.153
## F-statistic: 2.445 on 7 and 49 DF,  p-value: 0.03123
model_2 <- lm(data = combo_dat, final.performance ~ mean_log_testo + diversity_score + mean_log_testo*d
summary(model_2)

##
## Call:
## lm(formula = final.performance ~ mean_log_testo + diversity_score +
##     mean_log_testo * diversity_score + proportion_female + mean_age +
##     team.size + time.of.day + mean_log_cortisol, data = combo_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83246 -0.40210 -0.01763  0.45728  1.28480
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -37.86646    13.77647   -2.749  0.00841 **
## mean_log_testo     8.45597     2.83636    2.981  0.00450 **
## diversity_score    8.18115     2.88923    2.832  0.00675 **
## proportion_female  -0.69965     1.28435   -0.545  0.58845
## mean_age         -0.09484     0.12090   -0.784  0.43662
## team.size         0.48385     0.18153    2.665  0.01044 *
## time.of.day       0.06634     0.05115    1.297  0.20087
## mean_log_cortisol  -0.20664     0.23859   -0.866  0.39074
## mean_log_testo:diversity_score -1.85834     0.63418   -2.930  0.00517 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6557 on 48 degrees of freedom
## Multiple R-squared:  0.2703, Adjusted R-squared:  0.1486
## F-statistic: 2.222 on 8 and 48 DF,  p-value: 0.04199
```

```

anova(model_1, model_2)

## Analysis of Variance Table
##
## Model 1: final.performance ~ mean_log_testo + diversity_score + mean_log_testo *
##      diversity_score + proportion_female + mean_age + team.size +
##      time.of.day
## Model 2: final.performance ~ mean_log_testo + diversity_score + mean_log_testo *
##      diversity_score + proportion_female + mean_age + team.size +
##      time.of.day + mean_log_cortisol
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      49 20.962
## 2      48 20.640  1   0.32255 0.7501 0.3907

residual_1 <- resid(model_1)

```