ABCD Simulations for Manuscript

Kyle Baacke and Michael Nima Hekmat

2025-08-13

Kyle functions

Imports

Custom Functions

```
descriptives = function(df, vars, titles){
  ind = 1
  out = data.frame(
    mean=numeric(0),
    SD=numeric(0),
    SS=numeric(0),
    Var=numeric(0),
    Skewness=numeric(0),
    kurtosis=numeric(0),
    std_error=numeric(0),
    N=numeric(0)
  )
  for (v in vars){
    print(v)
    var = df[v]
    row_vector = c(
      mean(var[!is.na(var)]),
      sd(var[!is.na(var)]),
      sum((mean(var[!is.na(var)])-var[!is.na(var)])^2),
      var(var[!is.na(var)]),
      agricolae::skewness(var[!is.na(var)]),
      agricolae::kurtosis(var[!is.na(var)]),
      sd(var[!is.na(var)])/sqrt(length(!is.na(var))),
      length(var[!is.na(var)])
    )
    print(row_vector)
    out[ind,] = row_vector
    ind = ind+1
  }
  row.names(out) = titles
  print(format(out,scientific=F, digits=3))
  return(out)
}
```

```r
#https://www.statisticshowto.com/cohens-f-statistic-definition-formulas/
# Power: https://www.r-bloggers.com/2021/05/power-analysis-in-statistics-with-r/
#    pwr.f2.test {pwr}
summary_dfs = list()
detail_dfs = list()
lm_summary_dataframe = function(label_col_1, label_col_2, label_col_3, label_col_4, lm_object){
  # label_col_2, label_col_3, terms, Adjusted_R_Squared, F, df_1, df_2, p_value, residual_SE, n_terms
  summary_obj = summary(lm_object)
  res_data_frame = data.frame(
    label_col_1 = c(label_col_1),
    label_col_2 = c(label_col_2),
    label_col_3 = c(label_col_3),
    label_col_4 = c(label_col_4),
    terms = c(toString(lm_object$terms)),
    adjusted_r_squared = c(summary_obj$adj.r.squared),
    f = c(summary_obj$fstatistic['value']),
    df_1 = c(summary_obj$fstatistic['numdf']),
    df_2 = c(summary_obj$fstatistic['dendf']),
    p_value = c(pf(
      summary_obj$fstatistic['value'],
      summary_obj$fstatistic['numdf'],
      summary_obj$fstatistic['dendf'],
      lower.tail=F
      )),
    # residual_SE = c(), # TODO
    n_terms = c(length(names(lm_object$model))-1)
  )
  return(res_data_frame)
}

lm_detail_dataframe = function(label_col_1, label_col_2, label_col_3, label_col_4, lm_object){
  # subject_subset, variable_subset, variable_name, Estimate, Std_Error, t_value, p_value, VIF, beta, p
  coef_df = as.data.frame(summary(lm_object)$coefficients)
  # coef_df = coef_df[c(2:length(rownames(coef_df))),]
  coef_df$variable_name = rownames(coef_df)
  tryCatch(
    {
      coef_df$VIF = c(c(NA),vif(lm_object))
    },
    error=function(e){
      # print(e)
      coef_df$VIF = c(NA)
    }
  )
  tryCatch(
    {
      coef_df$Cohens_f2_partial_vec = c(NA,cohens_f_squared(lm_object)$Cohens_f2_partial)
      coef_df$Eta2_partial_vec = c(NA,eta_squared(lm_object)$Eta2_partial)
      coef_df$r2_partial_vec = c(NA,r2_semipartial(lm_object)$r2_semipartial)
    },
    error=function(e){
      coef_df$Cohens_f2_partial_vec = NA
      coef_df$Eta2_partial_vec = NA
```

```r
      coef_df$r2_partial_vec = NA
    }
  )

  power_vec = c()
  summary_obj = summary(lm_object)
  for (f_val in coef_df$Cohens_f2_partial_vec){
    print(f_val)
    tryCatch(
      {
        power = pwr.f2.test(u=summary_obj$df[1]-1, v=summary_obj$df[2], f2=f_val, sig.level=0.05)$power
      },
      error=function(e){
        power = c(NA)
      }
    )
    power_vec = c(power_vec, power)
  }
  coef_df$power = unlist(power_vec, use.names = FALSE)
  coef_df$label_col_1 = label_col_1
  coef_df$label_col_2 = label_col_2
  coef_df$label_col_3 = label_col_3
  coef_df$label_col_4 = label_col_4
  return(coef_df)
}
```

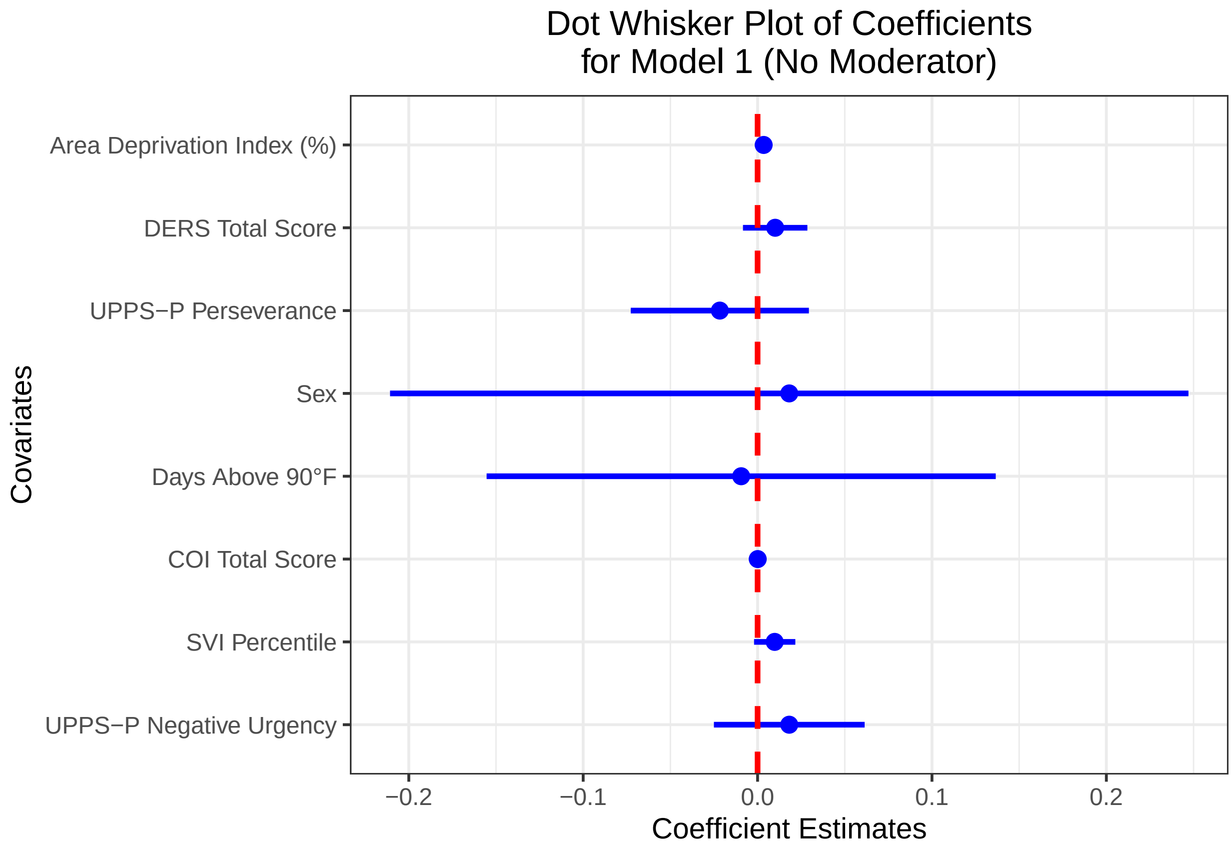
Model 1 observing effects of covariates without heat as a moderator

```r
summary(model_1)
```

```
##
## Call:
## lm(formula = mh_p_cbcl__synd__ext_sum ~ le_l_adi_addr1_national_prcnt +
##     sdev_y_ders_total + mh_y_upps_pers_sum + sex + Days_Above_90 +
##     le_l_coi_addr1_coi_total_national_score + le_l_svi_addr1_total_prcntile +
##     mh_y_upps_nurg_sum, data = abcd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4871  -4.0130  -0.0402   3.9501  23.3224
##
## Coefficients:
##                                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              3.943e+00  3.734e-01  10.560   <2e-16
## le_l_adi_addr1_national_prcnt            3.489e-03  2.171e-03   1.607    0.108
## sdev_y_ders_total                        1.009e-02  9.424e-03   1.071    0.284
## mh_y_upps_pers_sum                      -2.165e-02  2.606e-02  -0.831    0.406
## sex                                      1.816e-02  1.168e-01   0.156    0.876
## Days_Above_90                           -9.399e-03  7.447e-02  -0.126    0.900
## le_l_coi_addr1_coi_total_national_score  7.966e-05  2.139e-03   0.037    0.970
## le_l_svi_addr1_total_prcntile            9.807e-03  6.039e-03   1.624    0.104
```

```
## mh_y_upps_nurg_sum                        1.817e-02  2.205e-02   0.824    0.410
##
## (Intercept)                              ***
## le_l_adi_addr1_national_prcnt
## sdev_y_ders_total
## mh_y_upps_pers_sum
## sex
## Days_Above_90
## le_l_coi_addr1_coi_total_national_score
## le_l_svi_addr1_total_prcntile
## mh_y_upps_nurg_sum
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.838 on 9991 degrees of freedom
## Multiple R-squared:  0.0007732,  Adjusted R-squared:  -2.688e-05
## F-statistic: 0.9664 on 8 and 9991 DF,  p-value: 0.4602
```

```
dwplot(model_1,
       dot_args = list(color = "blue", size = 2.5),
       whisker_args = list(color="blue",size = 1)) %>%
  relabel_predictors(c(
    "(Intercept)" = "Intercept",
    "le_l_adi_addr1_national_prcnt" = "Area Deprivation Index (%)",
    "sdev_y_ders_total" = "DERS Total Score",
    "mh_y_upps_pers_sum" = "UPPS-P Perseverance",
    "sex" = "Sex",
    "Days_Above_90" = "Days Above 90°F",
    "le_l_coi_addr1_coi_total_national_score" = "COI Total Score",
    "le_l_svi_addr1_total_prcntile" = "SVI Percentile",
    "mh_y_upps_nurg_sum" = "UPPS-P Negative Urgency"
  )) +
  geom_vline(xintercept = 0, colour = "red", linetype = 2, linewidth = 1) +
  ggtitle(str_wrap("Dot Whisker Plot of Coefficients for Model 1 (No Moderator)", width =35)) +
  xlab("Coefficient Estimates") +
  ylab("Covariates") +
  theme_bw()+
  theme(plot.title = element_text(hjust = 0.5))
```



Model 1 Post-hoc and effects size (no moderator)

```r
# Eta squared
eta_squared(model_1, partial = TRUE)
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter                              | Eta2 (partial) |       95% CI
## ---------------------------------------------------------------------------
## le_l_adi_addr1_national_prcnt          |       2.56e-04 | [0.00, 1.00]
## sdev_y_ders_total                      |       1.12e-04 | [0.00, 1.00]
## mh_y_upps_pers_sum                     |       6.85e-05 | [0.00, 1.00]
## sex                                    |       1.77e-06 | [0.00, 1.00]
## Days_Above_90                          |       1.48e-06 | [0.00, 1.00]
## le_l_coi_addr1_coi_total_national_score |       4.53e-08 | [0.00, 1.00]
## le_l_svi_addr1_total_prcntile          |       2.65e-04 | [0.00, 1.00]
## mh_y_upps_nurg_sum                     |       6.79e-05 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```r
# Standardized betas
standardize_parameters(model_1)
```

```
## # Standardization method: refit
##
## Parameter                              | Std. Coef. |        95% CI
## -----------------------------------------------------------------
## (Intercept)                            |   6.08e-17 | [-0.02, 0.02]
## le l adi addr1 national prcnt          |       0.02 | [ 0.00, 0.04]
## sdev y ders total                      |       0.01 | [-0.01, 0.03]
## mh y upps pers sum                     |  -8.31e-03 | [-0.03, 0.01]
## sex                                    |   1.56e-03 | [-0.02, 0.02]
## Days Above 90                          |  -1.26e-03 | [-0.02, 0.02]
## le l coi addr1 coi total national score |   3.73e-04 | [-0.02, 0.02]
## le l svi addr1 total prcntile          |       0.02 | [ 0.00, 0.04]
## mh y upps nurg sum                     |   8.24e-03 | [-0.01, 0.03]
```

Model 2 observing effects of covariates with heat as a moderator

```r
summary(model_2)
```

```
##
## Call:
## lm(formula = mh_p_cbcl__synd__ext_sum ~ le_l_adi_addr1_national_prcnt *
##     Days_Above_90 + sdev_y_ders_total * Days_Above_90 + mh_y_upps_pers_sum *
##     Days_Above_90 + sex * Days_Above_90 + le_l_coi_addr1_coi_total_national_score *
##     Days_Above_90 + le_l_svi_addr1_total_prcntile * Days_Above_90 +
##     mh_y_upps_nurg_sum * Days_Above_90, data = abcd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4457  -3.9967  -0.0284   3.9621  23.3317
##
## Coefficients:
##                                                        Estimate Std. Error
## (Intercept)                                           3.7193710  0.4976331
## le_l_adi_addr1_national_prcnt                         0.0017343  0.0029323
## Days_Above_90                                         0.2978868  0.4820770
## sdev_y_ders_total                                     0.0162085  0.0126971
## mh_y_upps_pers_sum                                    0.0055157  0.0349103
## sex                                                   0.2139535  0.1566375
## le_l_coi_addr1_coi_total_national_score              -0.0004819  0.0028728
## le_l_svi_addr1_total_prcntile                         0.0175650  0.0080807
## mh_y_upps_nurg_sum                                    0.0072606  0.0294011
## le_l_adi_addr1_national_prcnt:Days_Above_90           0.0024964  0.0028062
## Days_Above_90:sdev_y_ders_total                      -0.0083859  0.0121240
## Days_Above_90:mh_y_upps_pers_sum                     -0.0381012  0.0334656
## Days_Above_90:sex                                    -0.2727332  0.1492309
## Days_Above_90:le_l_coi_addr1_coi_total_national_score  0.0006791  0.0027217
## Days_Above_90:le_l_svi_addr1_total_prcntile          -0.0107246  0.0077311
## Days_Above_90:mh_y_upps_nurg_sum                      0.0164876  0.0287513
##                                                      t value Pr(>|t|)
## (Intercept)                                            7.474 8.42e-14 ***
## le_l_adi_addr1_national_prcnt                          0.591   0.5542
## Days_Above_90                                          0.618   0.5366
```

```
## sdev_y_ders_total                                     1.277   0.2018
## mh_y_upps_pers_sum                                    0.158   0.8745
## sex                                                   1.366   0.1720
## le_l_coi_addr1_coi_total_national_score              -0.168   0.8668
## le_l_svi_addr1_total_prcntile                         2.174   0.0298 *
## mh_y_upps_nurg_sum                                    0.247   0.8050
## le_l_adi_addr1_national_prcnt:Days_Above_90           0.890   0.3737
## Days_Above_90:sdev_y_ders_total                      -0.692   0.4892
## Days_Above_90:mh_y_upps_pers_sum                     -1.139   0.2549
## Days_Above_90:sex                                    -1.828   0.0676 .
## Days_Above_90:le_l_coi_addr1_coi_total_national_score  0.250   0.8030
## Days_Above_90:le_l_svi_addr1_total_prcntile          -1.387   0.1654
## Days_Above_90:mh_y_upps_nurg_sum                      0.573   0.5663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.838 on 9984 degrees of freedom
## Multiple R-squared:  0.001622,   Adjusted R-squared:  0.0001222
## F-statistic: 1.081 on 15 and 9984 DF,  p-value: 0.3676
```

```r
interaction <- tidy(model_2) %>% filter(str_detect(term, ":"))
dwplot(interaction,
       dot_args = list(color = "blue", size = 3),
       whisker_args = list(color = "blue", size = 1.5)) %>%
  relabel_predictors(c(
    "le_l_adi_addr1_national_prcnt:Days_Above_90" = "ADI × Heat",
    "Days_Above_90:sdev_y_ders_total" = "DERS × Heat",
    "Days_Above_90:mh_y_upps_pers_sum" = "Perseverance × Heat",
    "Days_Above_90:sex" = "Sex × Heat",
    "Days_Above_90:le_l_coi_addr1_coi_total_national_score" = "COI × Heat",
    "Days_Above_90:le_l_svi_addr1_total_prcntile" = "SVI × Heat",
    "Days_Above_90:mh_y_upps_nurg_sum" = "Negative Urgency × Heat"
  )) +
  geom_vline(xintercept = 0, colour = "red", linetype = 2, linewidth = 1) +
  ggtitle(str_wrap("Dot Whisker Plot of Coefficients for Model 2 (With Moderator)", width = 35)) +
  xlab("Coefficient Estimates") +
  ylab("Covariates") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
```

Model 2 Post-hoc and effects size for moderation effects of heat

```r
# eta squared
eta_squared(model_2, partial = TRUE)
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter                                        | Eta2 (partial) |       95% CI
## --------------------------------------------------------------------------------
## le_l_adi_addr1_national_prcnt                    |       2.56e-04 | [0.00, 1.00]
## Days_Above_90                                    |       1.42e-06 | [0.00, 1.00]
## sdev_y_ders_total                                |       1.12e-04 | [0.00, 1.00]
## mh_y_upps_pers_sum                               |       6.85e-05 | [0.00, 1.00]
## sex                                              |       1.78e-06 | [0.00, 1.00]
## le_l_coi_addr1_coi_total_national_score          |       4.54e-08 | [0.00, 1.00]
## le_l_svi_addr1_total_prcntile                    |       2.66e-04 | [0.00, 1.00]
## mh_y_upps_nurg_sum                               |       6.80e-05 | [0.00, 1.00]
## le_l_adi_addr1_national_prcnt:Days_Above_90      |       8.28e-05 | [0.00, 1.00]
## Days_Above_90:sdev_y_ders_total                  |       5.46e-05 | [0.00, 1.00]
## Days_Above_90:mh_y_upps_pers_sum                 |       1.29e-04 | [0.00, 1.00]
## Days_Above_90:sex                                |       3.52e-04 | [0.00, 1.00]
## Days_Above_90:le_l_coi_addr1_coi_total_national_score |       6.96e-06 | [0.00, 1.00]
## Days_Above_90:le_l_svi_addr1_total_prcntile      |       1.91e-04 | [0.00, 1.00]
## Days_Above_90:mh_y_upps_nurg_sum                 |       3.29e-05 | [0.00, 1.00]
```

```
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```r
# standardized betas
standardize_parameters(model_2)
```

```
## # Standardization method: refit
##
## Parameter                                                | Std. Coef. |        95% CI
## ---------------------------------------------------------------------------------------
## (Intercept)                                              |   4.00e-04 | [-0.02, 0.02]
## le l adi addr1 national prcnt                            |       0.02 | [ 0.00, 0.04]
## Days Above 90                                            |  -7.11e-04 | [-0.02, 0.02]
## sdev y ders total                                        |       0.01 | [-0.01, 0.03]
## mh y upps pers sum                                       |  -8.12e-03 | [-0.03, 0.01]
## sex                                                      |   1.97e-03 | [-0.02, 0.02]
## le l coi addr1 coi total national score                  |  -3.04e-05 | [-0.02, 0.02]
## le l svi addr1 total prcntile                            |       0.02 | [ 0.00, 0.04]
## mh y upps nurg sum                                       |   8.53e-03 | [-0.01, 0.03]
## le l adi addr1 national prcnt × Days Above 90            |   9.02e-03 | [-0.01, 0.03]
## Days Above 90 × sdev y ders total                        |  -6.98e-03 | [-0.03, 0.01]
## Days Above 90 × mh y upps pers sum                       |      -0.01 | [-0.03, 0.01]
## Days Above 90 × sex                                      |      -0.02 | [-0.04, 0.00]
## Days Above 90 × le l coi addr1 coi total national score |   2.49e-03 | [-0.02, 0.02]
## Days Above 90 × le l svi addr1 total prcntile            |      -0.01 | [-0.03, 0.01]
## Days Above 90 × mh y upps nurg sum                       |   5.87e-03 | [-0.01, 0.03]
```

Finally... The differences in R^2...

```r
#R2
r2model1 = summary(model_1)$r.squared
r2model2 = summary(model_2)$r.squared

# adj R2
r2adjmodel1 = summary(model_1)$adj.r.squared
r2adjmodel2 = summary(model_2)$adj.r.squared

# data frame for comparison
df = data.frame(
  Models = c("Dot Whisker Plot of Coefficients for Model 2 (Moderator)"),
  R2 = c(r2model1, r2model2),
  AdjustedR2= c(r2adjmodel1, r2adjmodel2)
)
print(df)
```

```
##                                                     Models           R2
## 1 Dot Whisker Plot of Coefficients for Model 2 (Moderator) 0.0007732181
## 2 Dot Whisker Plot of Coefficients for Model 2 (Moderator) 0.0016221356
##      AdjustedR2
## 1 -2.688341e-05
## 2  1.221688e-04
```

Conclusions. . .

- **The model has low adjusted** R^2, with the moderator it increases a bit (still pretty low).
- **With moderation effects, Social Vulnerability Index is the only significant covariate.**
- **It may be interesting to see how this changes state by state since this model looks at things nationally.**