Assignment2

Alireza Aleali

## Data cleaning & sex recoding prior to analysis

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(haven)  
  
#importing data  
data = read\_sav("/Users/alireza/Desktop/Rclass/Assignment/2nd\_assignment/PSYR6003.A2.sav")  
head(data)

# A tibble: 6 × 21  
 RespondentId sex mps.SPP1.3y mps.SPP2.3y mps.SPP3.3y mps.SPP4.3y mps.SPP5.3y  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 4561355 Fema… 1 1 1 1 1  
2 4561556 Fema… 1 5 1 1 1  
3 4198369 Male 5 1 1 4 4  
4 4590305 Fema… 1 3 4 1 4  
5 4554366 Fema… 2 6 3 1 2  
6 4598695 Fema… 1 7 1 7 7  
# ℹ 14 more variables: guilt1.3y <dbl>, guilt2.3y <dbl>, guilt3.3y <dbl>,  
# dep1.3y <dbl>, dep2.3y <dbl>, dep3.3y <dbl>, fear1.3y <dbl>,  
# fear2.3y <dbl>, fear3.3y <dbl>, host1.3y <dbl>, host2.3y <dbl>,  
# host3.3y <dbl>, tipm.CONS1.3y <dbl>, tipm.CONS2.3y <dbl>

#Recoding 'sex' to a Factor  
recoded\_data =   
 data %>%   
 drop\_na() %>%  
 mutate(sex = as.factor(sex)) %>%   
 mutate(sex = recode\_factor(sex, Female = "0", Male = "1", "Other (specify)" = "2"))  
head(recoded\_data)

# A tibble: 6 × 21  
 RespondentId sex mps.SPP1.3y mps.SPP2.3y mps.SPP3.3y mps.SPP4.3y mps.SPP5.3y  
 <chr> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 4561355 0 1 1 1 1 1  
2 4561556 0 1 5 1 1 1  
3 4198369 1 5 1 1 4 4  
4 4590305 0 1 3 4 1 4  
5 4554366 0 2 6 3 1 2  
6 4598695 0 1 7 1 7 7  
# ℹ 14 more variables: guilt1.3y <dbl>, guilt2.3y <dbl>, guilt3.3y <dbl>,  
# dep1.3y <dbl>, dep2.3y <dbl>, dep3.3y <dbl>, fear1.3y <dbl>,  
# fear2.3y <dbl>, fear3.3y <dbl>, host1.3y <dbl>, host2.3y <dbl>,  
# host3.3y <dbl>, tipm.CONS1.3y <dbl>, tipm.CONS2.3y <dbl>

# Count NaN values in each column  
nan\_counts <- recoded\_data %>%   
 summarise(across(everything(), ~sum(is.nan(.))))  
  
# Print the count of NaN values  
print(nan\_counts)

# A tibble: 1 × 21  
 RespondentId sex mps.SPP1.3y mps.SPP2.3y mps.SPP3.3y mps.SPP4.3y mps.SPP5.3y  
 <int> <int> <int> <int> <int> <int> <int>  
1 0 0 0 0 0 0 0  
# ℹ 14 more variables: guilt1.3y <int>, guilt2.3y <int>, guilt3.3y <int>,  
# dep1.3y <int>, dep2.3y <int>, dep3.3y <int>, fear1.3y <int>,  
# fear2.3y <int>, fear3.3y <int>, host1.3y <int>, host2.3y <int>,  
# host3.3y <int>, tipm.CONS1.3y <int>, tipm.CONS2.3y <int>

## Calculating subscale totals & Reverse coding ‘tipm.CONS2.3y’

recoded\_data =   
 recoded\_data %>%   
 rowwise %>%   
 mutate(perf\_sc = mean(c(mps.SPP1.3y,mps.SPP2.3y,mps.SPP3.3y,mps.SPP4.3y, mps.SPP5.3y))) %>%   
 mutate(tipm.CONS2.3y = 8 - tipm.CONS2.3y) %>% #sub each from 8 since the scale is 1to7  
 mutate(conc\_sc = mean(c(tipm.CONS1.3y, tipm.CONS2.3y))) %>%   
 mutate(aff\_sc = mean(c(guilt1.3y, guilt2.3y, guilt3.3y,   
 dep1.3y, dep2.3y, dep3.3y,   
 fear1.3y, fear2.3y, fear3.3y,   
 host1.3y, host2.3y, host3.3y)))

## Hypotheses to Test

Previous studies indicate that both sex and conscientiousness are linked to negative affect. Specifically, females tend to experience more negative affect compared to males, and conscientiousness typically correlates negatively with negative affect. Additionally, socially prescribed perfectionism (SPP) tends to coincide with negative affect, wherein higher levels of SPP are associated with increased negative affect. However, it’s important to note that sex, conscientiousness, and SPP often vary together, making it challenging to identify the distinct effects of each factor. The aim of the current study is to demonstrate the additional predictive value of SPP in forecasting negative affect, beyond the influences of sex and conscientiousness.

**Hypothesis 1 (H1):** Gender, conscientiousness, and socially prescribed perfectionism (SPP) will collectively and significantly forecast negative affect, following the outlined patterns (i.e., women exhibiting higher negative affect than men, negative correlation between conscientiousness and negative affect).

**Research Question:** Are sex, conscientiousness, and SPP significant predictors of negative affect?

**Full model:** negative affect = b0(intercept) + b1(sex) + b2(conscientiousness) + b3(SPP) + e **Reduced model:** negative affect = b0

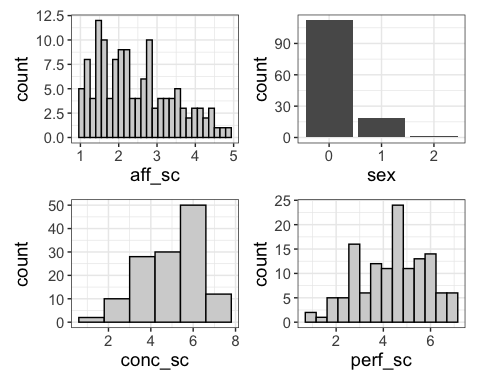
## Initial step involves visualizing the data through univariate plots.

library(flexplot)

Attaching package: 'flexplot'

The following object is masked from 'package:ggplot2':  
  
 flip\_data

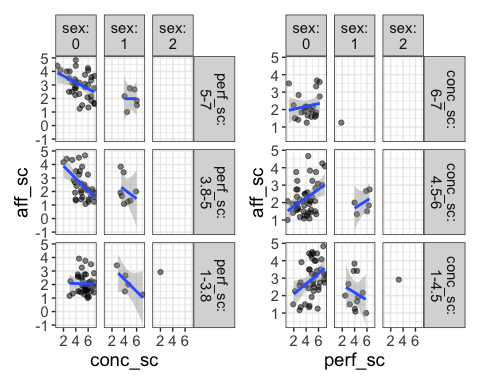
library(patchwork)  
aff\_uni =flexplot(aff\_sc~1, data = recoded\_data)  
sex\_uni =flexplot(sex~1, data = recoded\_data)  
conc\_uni =flexplot(conc\_sc~1, data = recoded\_data)  
perf\_uni =flexplot(perf\_sc~1, data = recoded\_data)  
aff\_uni+sex\_uni+conc\_uni+perf\_uni



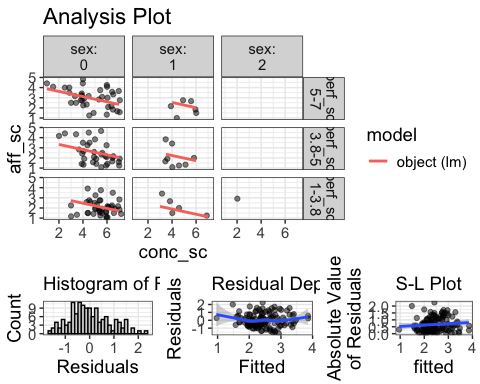
The scores on the negative affect subscale exhibit a slight right skew, however we can still consider almost a logical distribution for it. Regarding sex distribution, it’s unbalanced due to the presence of one individual identified as “other” (coded as 2), with fewer males than females. Conscientiousness demonstrates a peak around a mean value of approximately 6. Socially prescribed perfectionism (SPP) displays a distribution that is reasonably close to normal.

## Model visualization

#data visualization by considering sex distribution for other predictors  
fig1 = flexplot(aff\_sc ~ conc\_sc | sex+perf\_sc, data=recoded\_data, method = "lm")  
fig2 = flexplot(aff\_sc ~ perf\_sc | sex+conc\_sc, data=recoded\_data, method = "lm")  
fig1+fig2



#creating related model  
model1=lm(aff\_sc ~ sex + conc\_sc + perf\_sc, data=recoded\_data)  
visualize(model1)

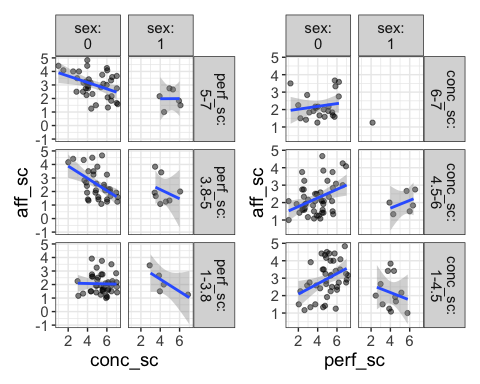


There exists a singular data point representing the “other” sex category. Additionally, there is only one data point corresponding to males with high conscientiousness, as observed through the binning process in flexplot. Among females (sex=0), a negative association between conscientiousness and negative affect appears apparent, with the strength of this relationship varying across different levels of SPP. A similar trend is observed among males (sex=1), albeit with fewer data points and variations observed across binned levels of perfectionism. Conversely, a positive correlation between SPP and negative affect is evident among females, with slight variations observed across different levels of conscientiousness. In males with lower and mid-levels of conscientiousness, the relationship between SPP and negative affect differs for each group, displaying a positive correlation in one and a negative correlation in the other.

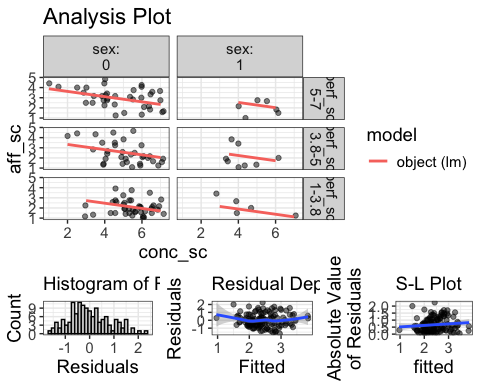
So, the data are independent since they are not longitudinal, and there are no dyads or groups of individuals. The histogram of residuals demonstrates a reasonably normal distribution, with a subtle right skew. However, the R-D plot exhibits curvature, indicating a potential violation of linearity. Additionally, the S-L plot displays a slight increase in slope, suggesting some degree of heteroskedasticity, although not severe.

## Conducting a sensitivity analysis by excluding one individual from the “other” sex category

recoded\_data2 =   
 recoded\_data %>%   
 filter(sex != "2")  
  
fig3= flexplot(aff\_sc ~ conc\_sc | sex+perf\_sc, data=recoded\_data2, method = "lm")  
fig4= flexplot(aff\_sc ~ perf\_sc | sex+conc\_sc, data=recoded\_data2, method = "lm")  
fig3+fig4



model2=lm(aff\_sc ~ sex + conc\_sc + perf\_sc, data=recoded\_data2)  
visualize(model2)



There is no particular difference! We need to check the model comparison to see which one is better

## Applying Model Comparison

# to see which model is better to use  
model.comparison(model1,model2)

Note: your models were fit to two different datasets.  
This is \*probably\* because you have missing data in one, but not the other.  
I'm going to make the dangerous assumption this is the case and do some ninja moves  
 in the background (hiya!). If you don't want me to do this, handle the missing data in advance

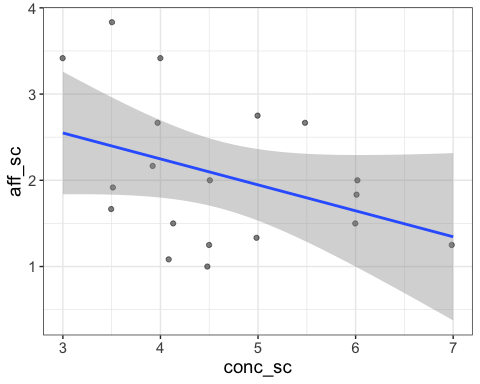
$statistics  
 aic bic bayes.factor rsq  
model1 333.257 347.633 1 0.269  
model2 333.257 347.633 1 0.269  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
 0 0 0 0 0

#estimates(model1)  
#estimates(model2)

From the model comparison we can see that there is no difference at all by considering or ignoring that individual data point (sex = other)

## Investigating male data to search for any possible outlier (since we found one single data point in higher level of conscientiousness)

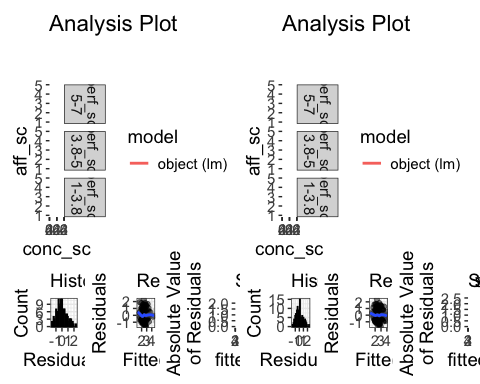
only\_male =   
 recoded\_data %>%   
 filter(sex==1)  
  
flexplot(aff\_sc ~ conc\_sc, data=only\_male, method = "lm")



It appears that the presence of the single “7” value does not significantly alter the relationship between conscientiousness and negative affect in males. Therefore, we can safely ignore to conduct a sensitivity analysis for this particular case.

## Since our model exhibits a violation of linearity, we can conduct sensitivity analyses by incorporating polynomial approach

model3 = lm(aff\_sc ~ sex + conc\_sc + perf\_sc + I(conc\_sc^2), data=recoded\_data)  
model4 = lm(aff\_sc ~ sex + conc\_sc + perf\_sc + I(perf\_sc^2), data=recoded\_data)  
visualize(model3) + visualize(model4)



#comparing each with original model which is "model1"  
  
#model1=lm(aff\_sc ~ sex + conc\_sc + perf\_sc, data=recoded\_data)  
  
model.comparison(model1,model3)

$statistics  
 aic bic bayes.factor p rsq  
model1 336.721 354.018 3.57 0.136 0.270  
model3 336.384 356.563 0.28 0.283  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.033 0.069 0.089 0.633

model.comparison(model1,model4)

$statistics  
 aic bic bayes.factor p rsq  
model1 336.721 354.018 2.136 0.074 0.270  
model4 335.357 355.536 0.468 0.289  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.038 0.089 0.119 0.477

model3&4 have slightly lower AIC (Akaike Information Criterion) than model1, suggesting that model3&4 might be slightly more efficient in terms of information loss. Despite model3&4 having a lower AIC, their BIC (Bayesian Information Criterion) is higher compared to model1. This implies that when considering the penalty for complexity, model1 might be preferred. model1 has a Bayes factor significantly higher than model3&4 ones. This strongly suggests that the evidence supports model1 over model3&4. Also, the p-values are above the common threshold for significance, suggesting that the models may not be explaining the variation in the data significantly better than a null model (model1). In addition, model3&4 have slightly higher R-squared values compared to model1, indicating that it explains more variance in the outcome. **Overall, model1 might be the more robust model despite its slightly lower R-squared value for using.**

## Checking for possible interaction

#The \* operator between sex, conc\_sc, and perf\_sc implies not just the inclusion of these variables but also all their interactions up to the three-way interaction (i.e., sex:conc\_sc, sex:perf\_sc, conc\_sc:perf\_sc, and sex:conc\_sc:perf\_sc).  
model5 = lm(aff\_sc ~ sex\*conc\_sc\*perf\_sc, data=recoded\_data)  
model.comparison(model1, model5)

$statistics  
 aic bic bayes.factor p rsq  
model1 336.721 354.018 4657.351 0.648 0.270  
model5 342.082 370.910 0.000 0.285  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.020 0.038 0.110 0.401

Still we like to proceed with **model1**!

Comparing full and reduced model based on our **hypothesis 1**

#full model = model1 = lm(aff\_sc ~ sex + conc\_sc + perf\_sc, data=recoded\_data)  
  
model\_reduced = lm(aff\_sc ~ 1, data=recoded\_data)  
model.comparison(model1, model\_reduced)

$statistics  
 aic bic bayes.factor p rsq  
model1 336.721 354.018 61624.36 <2e-16 0.27  
model\_reduced 370.310 376.076 0.00 0.00  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.010 0.164 0.335 0.545 1.487

estimates(model1)

Note: You didn't choose to plot conc\_sc so I am inputting the median

Note: You didn't choose to plot perf\_sc so I am inputting the median

Model R squared:  
0.27 (0.14, 0.4)  
  
Semi-Partial R squared:  
 sex conc\_sc perf\_sc   
 0.027 0.159 0.084   
  
Estimates for Factors:  
 variables levels estimate lower upper  
1 sex 0 2.54 2.38 2.69  
2 1 1.96 1.57 2.35  
3 2 2.39 0.67 4.11  
  
  
Mean Differences:  
 variables comparison difference lower upper cohens.d  
1 sex 1-0 -0.58 -1.28 0.13 -0.68  
2 2-0 -0.15 -2.99 2.70 -0.17  
3 2-1 0.43 -2.48 3.33 0.51  
  
  
Estimates for Numeric Variables =   
 variables estimate lower upper std.estimate std.lower std.upper  
1 (Intercept) 2.93 2.09 3.77 0.00 0.00 0.00  
2 conc\_sc -0.26 -0.37 -0.14 -0.35 -0.50 -0.19  
3 perf\_sc 0.20 0.10 0.31 0.30 0.15 0.45

Estimates for model comparison suggest that our model, including predictors, has the potential to forecast up to 1.5 points in negative affect. The selected model explains 27% of the variance in the outcome, with conscientiousness contributing the most (15.9%) to this explanation. Sex and SPP contribute 2.7% and 8.4%, respectively, to the variance explained by the model. All showing the superiority of **full model (model1)**

**Hypothesis 2 (H2):** Socially prescribed perfectionism (SPP) will forecast distinct variance in negative affect beyond the influences of sex and conscientiousness in a significant manner.

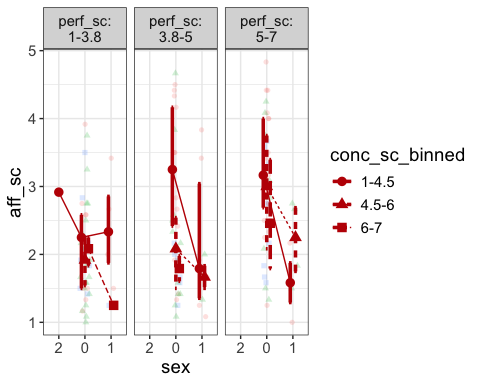
**Research Question:** Is SPP a significant predictor of negative affect, after controlling for the effects of gender and conscientiousness?

**Full model:** negative affect = b0 + b1(sex) + b2(conscientiousness) + b3(SPP) + e

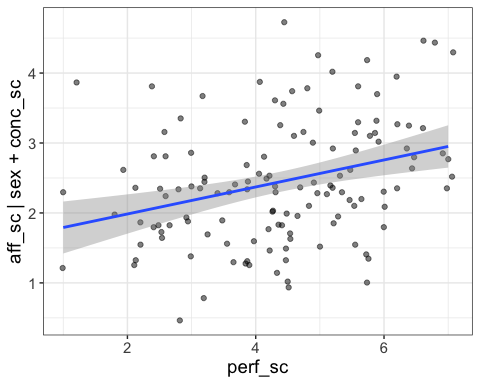
**Reduced model:** negative affect = b0 + b1(sex) + b2(conscientiousness) + e

Like previous one first we need visualization

flexplot(aff\_sc ~ sex + conc\_sc | perf\_sc, data=recoded\_data)



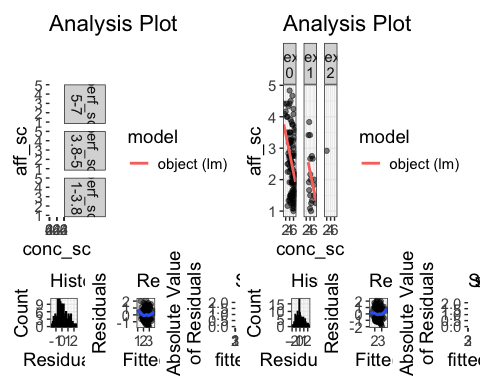
#since it's not really useful in terms of visualization we are switching to AVP  
added.plot(aff\_sc~perf\_sc, data = recoded\_data, method = "lm",  
 lm\_formula = aff\_sc~ sex + conc\_sc)



After controlling for conscientiousness and sex, there appears to be a positive association between SPP and negative affect. There is no necessity to recheck assumptions, as the model remains consistent with the one used for Hypothesis 1.

## Applying model comparison for full vs reduced models

#full model = model1 = lm(aff\_sc ~ sex + conc\_sc + perf\_sc, data=recoded\_data)  
model\_reduced\_2 = lm(aff\_sc ~ sex + conc\_sc, data=recoded\_data)  
visualize(model1) + visualize(model\_reduced\_2)



model.comparison(model1, model\_reduced\_2)

$statistics  
 aic bic bayes.factor p rsq  
model1 336.721 354.018 117.238 <2e-16 0.270  
model\_reduced\_2 349.132 363.546 0.009 0.186  
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.071 0.229 0.357 0.644

The **full model** can explain the predicted difference in negative affect of approximately 0.64 points. Comparatively, our complete model demonstrates superior fit, indicated by lower AIC/BIC values, a Bayes Factor exceeding 100, a significant p-value, and it accounts for 27% of the variance in the outcome, as opposed to only 18.6% explained by the reduced model.

## Descriptive Statistics: a. Bivariate correlation table

#creating correlation table including mean and SD  
correlation\_data = recoded\_data %>%   
 select(aff\_sc, sex, conc\_sc, perf\_sc)  
  
library(apaTables)  
table1 = apa.cor.table(correlation\_data, table.number=1, filename= "Correlation\_Table")  
  
print(table1)

Table 1   
  
Means, standard deviations, and correlations with confidence intervals  
   
  
 Variable M SD 1 2   
 1. aff\_sc 2.44 0.97   
   
 2. conc\_sc 5.01 1.31 -.37\*\*   
 [-.51, -.22]   
   
 3. perf\_sc 4.38 1.42 .36\*\* -.19\*   
 [.20, .50] [-.35, -.02]  
   
  
Note. M and SD are used to represent mean and standard deviation, respectively.  
Values in square brackets indicate the 95% confidence interval.  
The confidence interval is a plausible range of population correlations   
that could have caused the sample correlation (Cumming, 2014).  
 \* indicates p < .05. \*\* indicates p < .01.

table1 = as\_tibble(table1$table.body)  
write.csv(table1, file="/Users/alireza/Desktop/Rclass/Assignment/2nd\_assignment/table1.csv")

## b. Regression table

table2= apa.reg.table(model1, table.number = 2, filename = "Regression\_Table.doc")  
print(table2)

Table 2   
  
Regression results using aff\_sc as the criterion  
   
  
 Predictor b b\_95%\_CI sr2 sr2\_95%\_CI Fit  
 (Intercept) 2.93\*\* [2.08, 3.77]   
 sex1 -0.58\*\* [-0.99, -0.16] .04 [-.02, .10]   
 sex2 -0.15 [-1.87, 1.58] .00 [-.00, .00]   
 conc\_sc -0.26\*\* [-0.37, -0.14] .11 [.02, .20]   
 perf\_sc 0.20\*\* [0.10, 0.31] .08 [.00, .17]   
 R2 = .270\*\*  
 95% CI[.13,.37]  
   
  
Note. A significant b-weight indicates the semi-partial correlation is also significant.  
b represents unstandardized regression weights.   
sr2 represents the semi-partial correlation squared.  
Square brackets are used to enclose the lower and upper limits of a confidence interval.  
\* indicates p < .05. \*\* indicates p < .01.

table2 = as.tibble(table2$table\_body)

Warning: `as.tibble()` was deprecated in tibble 2.0.0.  
ℹ Please use `as\_tibble()` instead.  
ℹ The signature and semantics have changed, see `?as\_tibble`.

print(table2)

# A tibble: 8 × 6  
 Predictor b `b\_95%\_CI` sr2 `sr2\_95%\_CI` Fit   
 <chr> <chr> <chr> <chr> <chr> <chr>   
1 "(Intercept)" "2.93\*\*" "[2.08, 3.77]" "" "" ""   
2 "sex1" "-0.58\*\*" "[-0.99, -0.16]" ".04" "[-.02, .10]" ""   
3 "sex2" "-0.15" "[-1.87, 1.58]" ".00" "[-.00, .00]" ""   
4 "conc\_sc" "-0.26\*\*" "[-0.37, -0.14]" ".11" "[.02, .20]" ""   
5 "perf\_sc" "0.20\*\*" "[0.10, 0.31]" ".08" "[.00, .17]" ""   
6 "" "" "" "" "" "R2 = .270\*\*"   
7 "" "" "" "" "" "95% CI[.13,.37]"  
8 "" "" "" "" "" ""

write\_csv(table2, file="/Users/alireza/Desktop/Rclass/Assignment/2nd\_assignment/table2.csv")