



## DSC- 520 Multiple Linear Regression Analysis Guide

Follow the steps below to complete the assignment.

1. Create a new R Markdown file for this exercise. You will add the subsequent tasks to this file.
2. Create a synthetic dataset using the code provided in the "DSC-520 Multiple Regression – Dataset." Note that this snippet of R code is designed to generate data that approximates the data described in the article.
3. Inspect the core variables to assess their distribution (using the *testdistr* function) and identify outliers: MOOD, SSQ, SOLacti, DAS. Plot the distribution of each variable.
4. Examine the bivariate correlations between study variables: SSQ, MOOD, Stress, SOLacti, DAS, DBAS. Plot a heatmap depicting the correlations table (use the *plot* function and appropriate *theme*).
5. Create a basic table of graphs of descriptive statistics using the *egtable* function. Standardize the predictors to get standardized estimates (as in the article) using *as.vector(scale(variable\_name))*.
6. Fit three different models and compare them.
  - a. Model 1: Just the covariates
  - b. Model 2: Model 1 + main constructs of interest without interactions
  - c. Model 3: Model 2 + add the hypothesized interaction between subjective sleep quality and global dysfunctional beliefs
7. Combine the results of the three models into one table using the *screenreg()* function. Note the asterisk to the right of the threshold p-values and the errors in parentheses.
8. Since higher scores on subjective sleep quality indicate poorer sleep quality, assess which model (if any) shows that overall worse sleep quality and overall dysfunctional attitudes are significantly associated with more negative mood ( $p < .001$ ).
9. Ensure that the models are appropriate. Check the variance inflation factors (using the *vif* function) and the distribution of residuals (using the *testdistr* function). Describe and interpret your findings.
10. Refit the model on raw (i.e., nonstandardized data).
11. Use *ggplot* to visualize the relations between subjective sleep quality and negative mood, and the relation of poor sleep quality and mood of vulnerable adolescents with higher levels of dysfunctional attitudes. If you completed the previous steps correctly, the following R code should plot the graph:

```
ggplot(adosleep.newdat, aes(SSQ, MOOD, linetype=DAS)) +  
  geom_line(size = 2) +
```

```
scale_x_continuous("Subjective sleep quality\n(higher is worse)") +  
ylab("Negative Mood") +  
theme_cowplot() +  
theme(  
  legend.position = c(.85, .15),  
  legend.key.width = unit(2, "cm"))
```

12. Review the original objective of the analysis and ensure you were able to address the objectives, produce answers, and back up your claims with relevant calculations.