Name: Allama Ikbal Sijan

**Multivariate Statistics**

PSYC 739, Fall 2024

**Due Date**: September 30, 2024 (2:15 PM)

For this assignment, you will be using data from the **survey** dataset on Canvas. The variables for the assignment include:

* **sex:** participant’s sex (0 = Male, 1 = Female)
* **age:** participant’s age in years
* **age\_group**: a categorical version of the age variable based on younger adulthood, middle-aged participants, and older adulthood
* **child**: does the participant have a child (0 = No, 1 = Yes)
* **smoke**: is the participant a smoker (0 = No, 1 = Yes)
* **op1, Rop2, op3, Rop4, op5, Rop6:** 6 items assessing optimism [\*the negatively-worded items have already been reversed coded, so no need to reverse code any items yourself\*]
* **Mslfest, Mposaff, Mnegaff, Mpstress, Mlifesat**: mean scale scores of self-esteem, positive affect, negative affect, perceived stress, and life satisfaction

You will be completing the following questions, some of which involve conducting analyses in either SPSS or R (Parts A and B) and some analyses that need to be conducted in R (Part C). You will hand in the answers to the questions; enter them into this word document. Some of the answers involve you copying and pasting your SPSS syntax and/or R code. You will also hand in your *complete* SPSS syntax (i.e., the syntax for all the analyses that comprise this assignment; this is just good practice!) and your *complete* R code (saved in a script file).

***Homework #1, Part A: Introductory and Bivariate Statistics in SPSS or R***

1. Compute a scale score called **Optimism** using the six items (you can just use the mean function to create the score, i.e., method #2 on slide 7 in Session 2’s PowerPoint). Copy and paste your syntax here. This is your *optimism* variable. [4 points]

* df$Optimism <- rowMeans(df[ , c("op1", "Rop2", "op3", "Rop4", "op5", "Rop6")], na.rm = TRUE)

1. What is the Mean and Standard Deviation of **Optimism**? [2 points]

* *M* = 3.68, *SD* = 0.74

1. What is Cronbach’s alpha for the scale? Is this an acceptable level? Would alpha increase if any of the items were deleted? [4 points]

* Cronbach’s alpha = 0.8 for the optimism scale. Yes, this is an acceptable level.
* The alpha doesn’t increase but rather decreases if any of the items are deleted.

1. Create dummy variables based on the age\_group variable for each age group (i.e., **Younger\_Adults, MiddleAged\_Adults,** and **Older\_Adults**). Copy and paste your syntax here. [3 points]

* df <- df %>%
* mutate(MiddleAged\_Adults = if\_else(age\_group == 1, 1, 0)) %>%
* mutate(Older\_Adults = if\_else(age\_group == 2, 1, 0))

1. Now for some bivariate analyses! First, let’s see if there are differences in optimism between those with a child or not (variable “child” in the dataset). [5 points]
   1. What analysis should you do?
      1. t-test
   2. Write up the result of this analysis in APA format. (Note: this can be done in as short as one or two sentences.)
      1. There was a statistically significant difference in optimism between individuals with and without a child *t* (419) = -2.36, *p*  < .05. Those with a child (*M* = 3.78, *SD* = 0.68) reported higher levels of optimism than those without a child (*M* = 3.61, *SD* = 0.78).
2. Are there overall differences by age group in optimism? If so, what is the nature of these differences? (That is, which group[s] experienced more events than which other group[s]?) [\*Use the original “age\_group variable as the independent variable, not the dummy variables you created in Question 4\*] [7 points]
   1. What analysis should you do?
      1. One-Way ANOVA
   2. Write up the results of this analysis in APA format.
      1. I used one-way ANOVA to assess differences in optimism among younger adults, middle-aged adults, and older adults and found that there is a significant difference in optimism level by age group, *F* (2, 433) = 9.42, *p*  < .01. Pairwise comparisons demonstrated that older adult participants had significantly more optimism level than younger adult participants (Mdiff = -0.46, *p* < .01) and middle-aged participants (Mdiff = 0.29, *p*  = .01). Optimism did not significantly differ between middle-aged adults and younger adults (Mdiff = 0.16, p = .09)
3. Are there any differences between females and males in smoking status (variable “smoke” in the data set? [5 points]
   1. What analysis should you do?
      1. Chi-square (test of independence)
   2. Write up the results of this analysis in APA format.
      1. A chi-square test of independence showed no significant relationship between gender and smoking status in the sample, χ2(1) = 0.34, *p* = 0.56.
4. Create a correlation matrix table that includes the following variables: sex, age [\*use the scale version in the data set – “age”\*], smoking status, self-esteem, positive affect, negative affect, perceived stress, and life satisfaction.

No need to compute the scale scores for the psychosocial variables, just use the ones already included at the end of the data set (e.g., “Mslfest”, “Mposaff”, “Mnegaff”, “Mpstress”, and “Mlifesat”). Make sure to denote statistical significance with \* symbols as in the example from class. (Do not simply paste the SPSS/R default output table here but make sure an APA formatted table is included here.) [10 points]



***Homework #1, Part B: Multiple Regression in SPSS or R***

You will be using the same dataset as Part A for these additional questions. Keep saving your SPSS syntax and/or R code in the same file.

1. Run a **standard multiple regression** predicting negative affect levels (“Mnegaff”). Enter sex, child, smoking status, and perceived stress (“Mpstress”) as predictors. Remember to save your syntax. Enter the following values (and, if applicable, standard error and/or significance level). For each, also provide a brief explanation/interpretation of what the finding means. [10 points]

|  |  |  |
| --- | --- | --- |
| Statistic | Value | Explanation/Interpretation |
| *R2* | 0.46 | 46% of the variance in negative affect levels are explained by the sets of IVs. |
| *F* | 90.12, *p*  < .001 | The large F value with *p* < .001 indicates that our model is effective at explaining the variance in the DV (negative affect levels). |
| Constant | -0.28, *SE*  = 0.13, *p*  < .05 | Predicted negative affect levels when all the IVs are equal to zero. Meaning it’s the predicted negative affect levels of non-smoker male without a child with a perceived stress level of 0. |
| *B* for sex | -0.004, *SE*  = 0.05, *p*  = .93 | Being female (1) is associated with a 0.004 lower negative affect levels than being male when other IVs are held constant. But the association is not significant (*p*  = .93) |
| *B* for perceived stress | 0.83, *SE*  = 0.04, *p* < .001 | 1 unit increase in perceived stress is associated with a 0.83-unit increase in negative affect when other IVs are held constant. |

1. Next, you will be running a **hierarchical regression** predicting negative affect levels (“Mnegaff”), using the same variables as Question #9. How many steps will the analysis have, and which variables will you enter in each step? Why did you organize your variables this way? (Note: there is not a single “correct” answer here, just explain your reasoning). [10 points]

* To predict negative affect levels using a hierarchical regression, a total of 4 steps are required. I organized the variables based on the inverse of their predictive power (beta coefficients) of the DV and logical relevance. In the first step, the demographic variable sex was added as the sole IV as it was the least associated IV out of the 4. In the second step, the variable “smoke” was entered into the regression model, the second least associated IV. In step three, the variable the variable "child" was entered, considering that having children can influence affect both positively and negatively. Finally, in the fourth step, perceived stress was added, as it is highly likely to be associated with negative affect.

1. Run your hierarchical regression analysis. Remember to save your syntax. Make a table (in APA format) that summarizes the results for your individual predictors across the model steps. [10 points]

Table 2. Results of Hierarchical Regression Model Predicting Negative Affect (*N* = 436*)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Model 1** | | **Model 2** | | **Model 3** | | **Model 4** | |
|  | *B* | *SE* | *B* | *SE* | *B* | *SE* | *B* | *SE* |
| Constant | 1.87\*\*\* | 0.05 | 1.86\*\*\* | 0.05 | 1.90\*\*\* | 0.06 | -0.27\* | 0.13 |
| Sex |  |  |  |  |  |  |  |  |
| Male | -- | -- | -- | -- | -- | -- | -- | -- |
| Female | 0.12 | 0.07 | 0.13 | 0.07 | 0.13 | 0.07 | -0.004 | 0.05 |
| Smoke |  |  |  |  |  |  |  |  |
| No |  |  | -- | -- | -- | -- | -- | -- |
| Yes |  |  | 0.06 | 0.09 | 0.05 | 0.09 | 0.03 | 0.07 |
| Child |  |  |  |  |  |  |  |  |
| No |  |  |  |  | -- | -- | -- | -- |
| Yes |  |  |  |  | -0.09 | 0.07 | -0.04 | 0.05 |
| Perceived Stress |  |  |  |  |  |  | 0.83\*\*\* | 0.04 |

\* *p* < .05 \*\* *p* < .01 \*\*\* *p* < .001

1. Lastly, write up a Data Analysis Plan and Results for your hierarchical regression. For the Data Analysis Plan, only include the hierarchical regression (i.e., no need to include any preliminary analyses). For the Results, make sure to include results for your overall model at each step, as well as for the individual predictors. You can refer to your table from Question #11. Do your best to adhere to APA formatting. [10 points]

* Data Analysis Plan: A hierarchical regression analysis was performed to predict negative affect levels and determine if additional IVs significantly improve the prediction after accounting for the effects of previously entered IVs. Model 1 included only perceived stress as the IV. In Model 2, the variable "child" (0 = No, 1 = Yes) was entered. Model 3 introduced the variable "smoke" (0 = No, 1 = Yes). Finally, in Model 4 the variable "sex”(0 = Male, 1 = Female) was entered.
* Results: The results of the hierarchical regression analysis predicting negative affect are shown in Table 2. Model 1 with sex as the sole IV was not statistically significant [*F*(1, 434) = 3.16, *p* = .08]. Thus, it indicates sex was not significantly associated with negative affect levels. Model 2 with “smoke” as an added predictor was also not statistically significant [*F*(1, 430) = 1.91, *p* = .15]. Thus, it indicates smoking habit was not significantly associated with negative affect levels. Model 3 with “child” as an added predictor was also not statistically significant [*F*(1, 430) = 1.82, *p* = .14]. Thus, it indicates having children or not was not significantly associated with negative affect levels. Model 4 with perceived stress as an added predictor was statistically significant [*F*(1, 430) = 90.12, *p* <.001]. Thus, it indicates that perceived stress was significantly associated with negative affect levels.

***Homework #1, Part C: R!!!***

For this part of the assignment, load the **survey.csv** file into RStudio. The variables are the same as the example above. Make sure the first row of the data set is treated as headers when importing the file into RStudio.

1. Use the describe() function from the **psych** package to get the descriptive statistics for the Positive Affect variable (Mposaff). Copy/paste the R code below. [2 points]

* describe(df$Mposaff)

1. What is the Mean and Standard Deviation of the Positive Affect variable? [2 points]

* *M*  = 3.37, *SD*  = 0.73

1. Use the table() function to run a frequency analysis of the child variable. Copy/paste the R code below. [2 points]

* table(df$child)

1. How many participants do not have a child and how many participants do have a child? [2 points]

* No Child = 253
* Child = 185

1. Run an independent samples t-test with Positive Affect as the dependent variable and child as the independent variable. Assume that the variances in the two groups are equal. Copy/paste the R code below. Were the two groups significantly different in Positive Affect? If so, which group was significantly higher or lower? [2 points]

* t.test(df$Mposaff ~ df$child, var.equal = TRUE)
* There was a statistically significant difference in positive affect between individuals with and without a child *t* (433) = -2.36, *p* < .05. Those with a child (M = 3.46, SD = 0.72) reported higher levels of positive affect than those without a child (M = 3.30, SD = 0.73).

1. Run a Pearson’s correlation between Life Satisfaction and Perceived Stress. Copy/paste the R code below. What was the Pearson’s correlation coefficient? [2 points]

* corr.test(df$Mlifesat, df$Mpstress)
* -0.49

1. Perform a multiple linear regression with Self-Esteem as the dependent variable and age [the scale version of the variable], Smoking Status, and Perceived Stress as the independent variables. Copy/paste your R code below. [3 points]

* m\_model <- lm(Mslfest ~ age + smoke + Mpstress, data = df)
* summary(m\_model)

1. Use the summary() function to find the following values for the regression model in Question #7 [3 points]:
   1. R-squared: 0.35
   2. Adjusted R-squared: 0.34
   3. **F-statistic**: *F*(3, 424) = **74.51,** *p* < .001
2. Which independent variables (if any) were significant predictors of Self-Esteem in the regression model in Question #7? [2 points]

* Age (*B* = 0.003, *p <* .05) and Perceived Stress (*B* = -0.53, *p <* .001) were significant predictors of Self-Esteem in the model.