**Multiple Regression Lab**

**For the following lab worksheet please open the regress.sav file on Moodle.**

Your friend just finished his research study examining what factors predict the number of physical health symptoms. Your friend wrote the following write up about the number of stressful life events (stress) on the number of physical health symptoms (phyheal)*.*

*“A linear regression analysis found that the number of stressful life events could significantly predict the number of physical health symptoms one has, F(1, 463) = 47.678, p < .001. The number of stressful life events causes 9.1% of the explained variance in the number of physical health symptoms “*

**Q1: Even before looking at his data file, what is the major problem with your friend’s write up above? Multivariate Analysis does not indicate causality.**

**Q2: How would you interpret the effect size for the predictability of physical health symptoms on mental health visits? Moderate. R squared rule of thumb – Small = 1%, Moderate = 9%, Strong = 25%**

Because of what you have been taught about regression in Stats I and II you know that there may be some underlying issues with your friend’s interpretation of the results. After asking to see his dataset you notice that two more variables were not used in your friend’s original regression model; visits to health professionals (timedrs) and the number of mental health symptoms (menhel). Being the good friend that you are, you decide to run an atheoretical regression analysis on your friend’s data.

*Instructions: run a standard multiple regression analysis on all the data. Be sure that you use physical health symptoms as the dependent variable and answer the next several questions.*

**Q3. Is there any evidence that suggests violations of assumption for multiple regressions analysis (eg Independence of Observations, linearity, homoscedasticity, multicollinearity, outlier, normality, etc…)? Please explain (write in APA format).**

**There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.957. There was linearity and homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by VIF values greater than 10 thus no transformation of variables. There were 6 studentized deleted residuals greater than ±3 standard deviations, however, these cases remained in the final analysis due to no leverage values greater than 0.2, and values for Cook's distance above 1.** **The assumption of normality was met, as assessed by a Q-Q Plot.**

**Q4. Examine the model summary and fit. In APA format, explain if the model is a good fit and how much variance it explains if it is.**

**The multiple regression model statistically significantly predicted physical health symptoms, F(3,461) = 87.107, p < .001, adj. R2 = .36.**

**Q5. Summarize the coefficients, standard error, and standardized coefficient in an APA formatted summary table (see Standard Multiple Regression, pg 19 in Laerd for example).**

**Table 1**

Multiple Regression results for physical health symptoms

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| phyheal | *B* | 95% CI for B | | *SE B* | *β* | *R2* | *Adj R2* |
|  |  | *LL* | *UL* |  |  |  |  |
| Model |  |  |  |  |  | .362 | .358\*\*\* |
| Constant | 2.79\*\*\* | 2.43 | 3.15 |  |  |  |  |
| Visits to health Professionals | .069\*\*\* | .053 | .086 | .009 | .318\*\*\* |  |  |
| Mental Health Symptoms | .227\*\*\* | .182 | .213 | .023 | .399\*\*\* |  |  |
| Stressful life events | .001 | .000 | .003 | .001 | .067 |  |  |

Note. Model = “Enter” method in SPSS Statistics.

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

**Q6. From the table above, interpret your coefficients. Please write the multiple regression equation.**

**Phyheal = 2.793 + .069(timedrs) + .227(menheal) + .001(stress)**

**Q7. Are there any factors that may not be good predictors in the model? Which one and how do you know it is not a good fit?**

**Stressful life events were not a good predictor of physical health symptoms because p>.05.**

After you ran the standard multiple regression model you noticed that one of the factors may not be a good predictor in the model. You have a theory that by itself, the factor in question is a significant predictor of physical health symptoms but as you add other factors it loses its power.

*Instructions: Run a hierarchical multiple regression analysis on all the data with physical health symptoms as the dependent variable. Step 1 please use only stress. Step 2 add timedrs. Step 3 add menheal. Please answer the following questions.*

**Q8. Examine the model summary and fit. In APA format, explain if the model is a good fit and how much variance it explains if it is. Please note how R squared change as you add the multiple factors**

**The full model of stressful events, visits to health professionals and mental health symptoms to predict physical health symptoms (PHS) (Model 3) was statistically significant, R2 = .362, F(3, 461) = 87.107, p < .0005; adjusted R2 = .358. The initial model for the prediction of (PHS) led to a statistically significant increase in R2 of .093, F(1, 463) = 47.678, p< .001. The addition of visits to health professionals to the prediction of PHS (Model 2) led to a statistically significant increase in R2 of .135, F(1,462) = 80.800, p < .001. The addition of mental health symptoms to the prediction of PHS (Model 3) also led to a statistically significant increase in R2 of .133, F(1, 461) = 96.399, p < .001.**

**Q10. Summarize the coefficients, standard error, and standardized coefficient in an APA formatted summary table (see Hierarchical Multiple Regression, pg 11 in Laerd for example).**

**Table 2**

*Hierarchical Multiple Regression Predicting Physical Health Symptoms*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Physical Health Symptoms | | | | | |
|  | Model 1 | | Model 2 | | Model 3 | |
| Variable | *B* | *β* | *B* | *β* | *B* | *β* |
| Constant | 3.875\*\*\* |  | 3.608\*\*\* |  | 2.793\*\*\* |  |
| Stressful Life Events | .005\*\*\* | .306 | .003\*\*\* | .196 | .001 | .067 |
| Visits to health  Professionals |  |  | .084\*\*\* | .383 | .069\*\*\* | .318 |
| Mental Health Symptoms |  |  |  |  | .227\*\*\* | .399 |
|  |  |  |  |  |  |  |
| *R2* | .093 |  | .228 |  | .362 |  |
| *F* | 47.678\*\*\* |  | 68.348\*\*\* |  | 87.107\*\*\* |  |
| *R2 Change* | .093 |  | .135 |  | .133 |  |
| *Change F* | 47.678\*\*\* |  | 80.80\*\*\* |  | 95.399\*\*\* |  |

Note. Model = “Enter” method in SPSS Statistics.

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

**Q11. What happens to the predictability of stressful life events on physical health symptoms as you add more factors? Which factor added do you think impacted the predictability of stressful life events the most?**

**Stressful life events no longer are a statistically significant predictor after adding mental health symptoms in Model 3.**

Now you are interested in what is the model equation to predict physical health symptoms. You decided to use a more exploratory procedure.

*Instructions: Run a* ***stepwise*** *multiple regression analysis on all the data with physical health symptoms as the dependent variable. Please answer the following questions.*

**Q12 Examine the model summary and fit. In APA format, explain if the model is a good fit and how much variance it explains if it is. Please note how R squared changes as you add the multiple factors (see Hierarchical Multiple Regression, pg 11 in Laerd for example).**

**The full model of visits to health professionals and mental health symptoms to predict physical health symptoms (PHS) (Model 2) was statistically significant, R2 = .358, F(2, 462) = 128.87, p < .0005; adjusted R2 = .355. The addition of visits to health professionals to the prediction of PHS (Model 2) led to a statistically significant increase in R2 of .103, F(1,462) = 74.230, p < .0005.**

**Q13. Summarize the coefficients, standard error, and standardized coefficient in an APA formatted summary table (see Hierarchical Multiple Regression, pg 11 in Laerd for example).**

**Table 3**

*Stepwise Multiple Regression Predicting Physical Health Symptoms*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Physical Health Symptoms | | | | | |
|  | Model 1 | | Model 2 | | Model 3 | |
| Variable | *B* | *β* | *B* | *β* | *B* | *β* |
| Constant | 3.211\*\*\* |  | 2.935\*\*\* |  |  |  |
| Mental Health Symptoms | .288\*\*\* | .505 | .239\*\*\* | .420 |  |  |
| Visits to health Professionals |  |  | .072\*\*\* | .332 |  |  |
| Stressful Life Events |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *R2* | .255 |  | .358 |  |  |  |
| *F* | 158.45\*\*\* |  | 128.87\*\*\* |  |  |  |
| *R2 Change* | .255 |  | .103 |  |  |  |
| *Change F* | 158.45\*\*\* |  | 74.23\*\*\* |  |  |  |

Note. A third model was found to be unnecessary in a stepwise multiple regression analysis

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

**Q14. Examine the variance of the variance explained from model 2 of your stepwise multiple regression and compare it to the variance explained from the standard multiple regression analysis you did earlier. What is the difference in variance between the two models? Why do you think the stepwise regression suggests a two-factor model versus the three-factor model in the standard regression?**

**The variance, though similar, is slightly higher for the three-factor model (R2 = .362) than the two-factor model (.358). The stepwise model suggested a two-factor model because stressful life events statistically did not add significantly to the model.**

**Q15. Of the two factors from the second model, which one is the better predictor of physical health symptoms? How do you know?**

**Mental health symptoms because the Beta Standardized Coefficient (β =.420) is higher than visits to health professionals (β =.332).**

**You are done! Please save your work and all your files. Save your output as MultiReg (LastName).**