Multiple Regression

Rosalie Valdres

TWU

Multiple Regression

Assign 3: Multiple Regression

Use the sleep5ED.sav data set and the multiple correlation/regression procedures that you have learned to use in SPSS to generate the statistics necessary to complete the following tasks. Place your answers in a word file and post in the assignment folder labeled *regression* under the **module 3 assessments** tab by **November 29.**

1. **Run a standard multiple regression procedure to explore how factors such as gender (sex), age (age), physical fitness (firate) and depression (depress) impact the level of sleepiness and associated symptoms (*totSAS)*.**

```
[DataSet1] C:\Users\rvaldres\Desktop\Valdres.sleep5ED.new.sav..sav
```

Descriptive Statistics

	Mean	Std. Deviation	N
sleepy & assoc sensations scale	26.0451	10.52051	222
sex	.45	.498	271
age	43.87	12.684	248
HADS Depression	3.50	2.993	269
physical fitness	6.42	1.717	266

Correlations

		sleepy & assoc sensations scale	sex	age	HADS Depression	physical fitness
Pearson Correlation	sleepy & assoc sensations scale	1.000	-.199	-.141	.482	-.267
	sex	-.199	1.000	-.017	-.071	.110
	age	-.141	-.017	1.000	-.004	-.039
	HADS Depression	.482	-.071	-.004	1.000	-.314
	physical fitness	-.267	.110	-.039	-.314	1.000

Sig. (1-tailed)	sleepy & assoc sensations scale	.	.001	.017	.000	.000
	sex	.001	.	.393	.124	.037
	age	.017	.393	.	.473	.271
	HADS Depression	.000	.124	.473	.	.000
	physical fitness	.000	.037	.271	.000	.
N	sleepy & assoc sensations scale	251	251	230	249	247
	sex	251	271	248	269	266
	age	230	248	248	246	243
	HADS Depression	249	269	246	269	265
	physical fitness	247	266	243	265	266

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	physical fitness, age, sex, HADS Depression[b]	.	Enter

a. Dependent Variable: sleepy & assoc sensations scale

b. All requested variables entered.

a. State 2 applicable research questions (Hint: use questions on page 154 of SPSS survival manual as an example).

Utilizing the example questions on page 154 of SPSS survival manual, this is how I wrote my question:

- How well the variables: age, gender, fitrate and depress predict the level of sleepiness and associated symptoms (totsas)?

- Which variable best predict totsas?

I use the standard multiple regression statistical analysis on page 154 & 155 SPSS manual. The results generated descriptive statistics, coefficients, casewise diagnostics, residual statistics, p –plot and scatter plot which showed which variable best predict the level of sleepiness and associated symptoms (totsas) which answered my 2 questions.

b. **Were the assumptions met? Was the sample adequate? Did you spot any multicollinearity? Address the checks used for normality, homoscedasticity and linearity? Did you discover any outliers? If there were problems what did you decide to do? Provide relevant computer output and rationale to support your checks of the assumptions**.

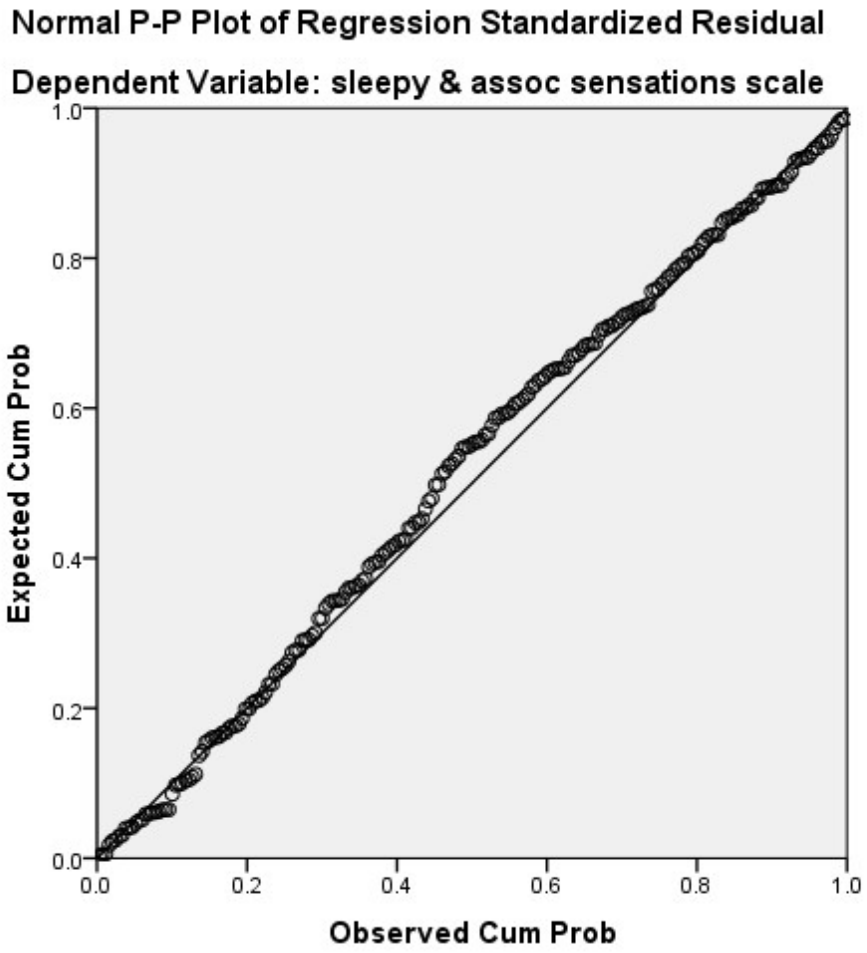
- The Assumptions were met in this data analysis. To check the assumptions: there must be multicollinearity, outliers, normality, linearity, homoscedasticity, independence of residuals and sample size. According to statistical solutions, in linear regression the sample size rule of thumb is that the regression analysis requires at least 20 cases per independent variable analysis. In this data, there are 251 cases, which met the rule.

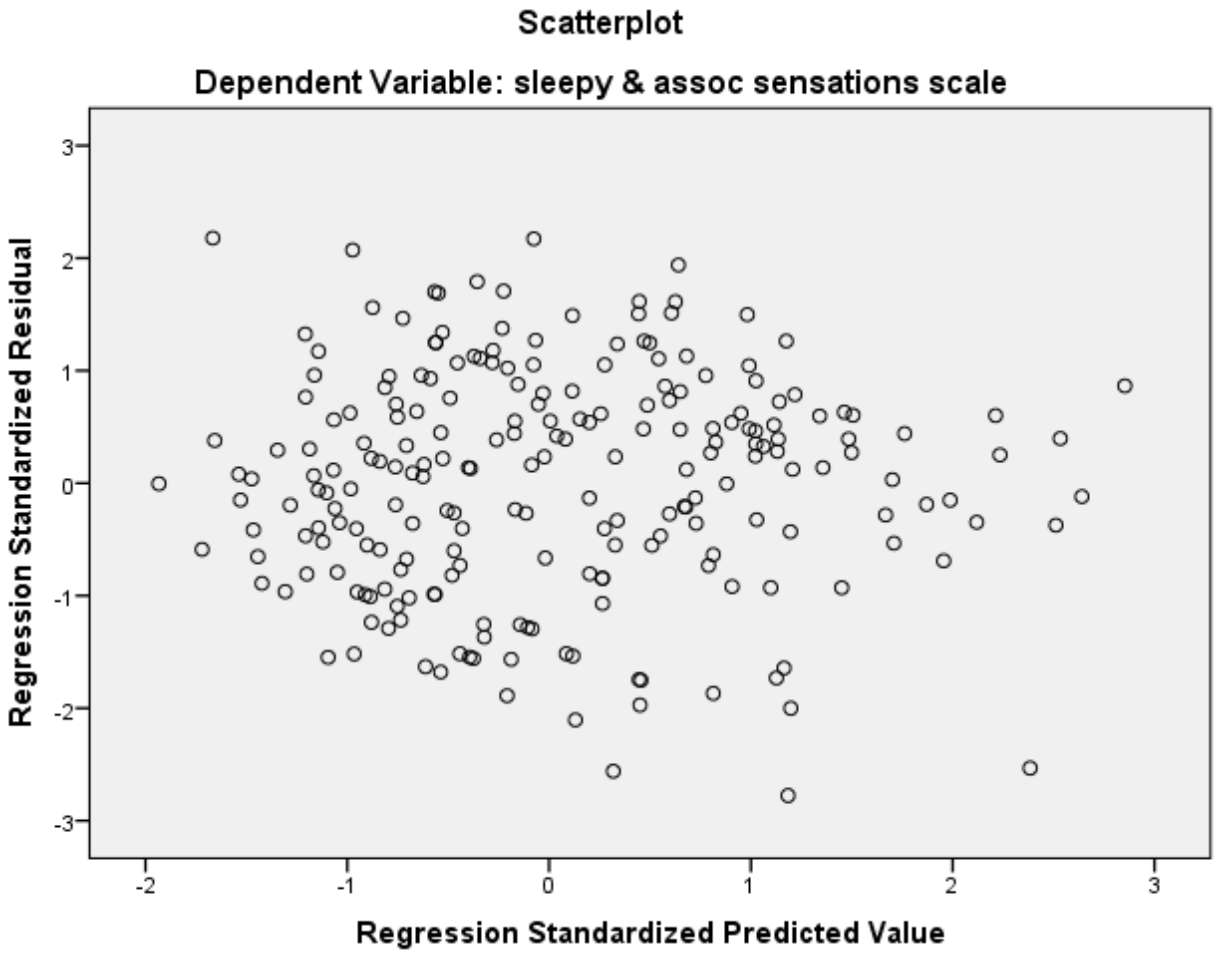
Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	sex	.986	1.014
	age	.998	1.002
	HADS Depression	.900	1.111
	physical fitness	.892	1.121

a. Dependent Variable: sleepy & assoc sensations scale

Under the coefficients table, I can check the multicollinearity but look at the VIF results on all four variables. A VIF value of above ten would be a concern, indicating multicollinearity (Pallant, 2016, page 159). I did not see multicollinearity in this analysis since all the variables have VIF results ranging from 1.002 to 1.121.


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: sleepy & assoc sensations scale
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0



- As for the normal P-Plot of the regression standardized residual, the points lie in a reasonably straight diagonal line from left to the top right which suggests no significant deviations from normality.
- In the scatterplot of the standardized residuals, it is rectangularly distributed with most of the scores concentrated at the center. Deviation from this suggests a violation of the assumption. There is no deviation in this data
- The presence of outliers can be detected from the scatterplot. Outliers can also be checked by inspecting the Mahalanobis Distance produced by the multiple regression program. It is found at the end of the file Mah_1
- Procedure: Go to Transform -> compute>I name under target variable probability_MD > Numeric expression box (1-chisq Probable_MD 1-3) click

OK. Then go to variable view look at the decimals – probable _MD, change to 5. If the probability_MD is <.001 this is considered an outlier. The result of my run ranges from .00318 to .82091 – which suggests that there are no outliers (Pallant, 2016, page 160)

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	15.04	42.28	26.152	5.6534	242
Std. Predicted Value	-1.933	2.853	.018	.9931	242
Standard Error of Predicted Value	.816	2.269	1.281	.2834	242
Adjusted Predicted Value	15.04	41.95	26.275	5.5775	225
Residual	-24.781	19.438	.178	8.8391	225
Std. Residual	-2.776	2.178	.020	.9901	225
Stud. Residual	-2.810	2.207	.020	1.0021	225
Deleted Residual	-25.393	19.980	.172	9.0501	225
Stud. Deleted Residual	-2.854	2.226	.019	1.0061	225
Mahal. Distance	.919	13.804	3.949	2.3142	242
Cook's Distance	.000	.075	.005	.0095	225
Centered Leverage Value	.004	.060	.017	.0104	242

a. Dependent Variable: sleepy & assoc sensations scale

- The Mahalanobis Distance results from this table showed a mean of 3.949, SD of 2.314 with a minimum of .919 and maximum of 13.804.

- Casewise Diagnostics table –give us information concerning unusual cases. This presents information about cases that have standardized value above 3.0 or below -3.0. This data does not show any of these numbers.

- Cook's Distance is used to check this strange case is having any undue influence on the results of our model as a whole. According to Tabachnick and Fidell (2013, page 75), cases with values larger than 1 are a potential problem. The table shows no value greater than one. (Actual results ranges from .000 to .075). If there is a value greater than 1; then this will be removed. (Pallant, 2016, page 161)

c. Present and interpret your findings. Copy and paste the relevant output from your SPSS run into your word file and provide a narrative summary of the output. Be sure to answer your research questions.

2. **Correlations**

		sleepy & assoc sensations scale	sex	age	HADS Depression
Pearson Correlation	sleepy & assoc sensations scale	1.000	-.199	-.141	.482
	sex	-.199	1.000	-.017	-.071
	age	-.141	-.017	1.000	-.004
	HADS Depression	.482	-.071	-.004	1.000
	physical fitness	-.267	.110	-.039	-.314
Sig. (1-tailed)	sleepy & assoc sensations scale	.	.001	.017	.000
	sex	.001	.	.393	.124

	age	.017	.393	.	.473
	HADS Depression	.000	.124	.473	.
	physical fitness	.000	.037	.271	.000
N	sleepy & assoc sensations scale	251	251	230	249
	sex	251	251	248	269
	age	230	248	248	246
	HADS Depression	249	269	246	269
	physical fitness	247	266	243	265

Correlations

		physical fitness
Pearson Correlation	sleepy & assoc sensations scale	-.267
	sex	.110
	age	-.039
	HADS Depression	-.314
	physical fitness	1.000
Sig. (1-tailed)	sleepy & assoc sensations scale	.000
	sex	.037
	age	.271
	HADS Depression	.000
	physical fitness	.
N	sleepy & assoc sensations scale	247
	sex	266
	age	243
	HADS Depression	265
	physical fitness	266

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.541[a]	.293	.280	8.927

a. Predictors: (Constant), physical fitness, age, sex, HADS Depression

b. Dependent Variable: sleepy & assoc sensations scale

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7413.343	4	1853.336	23.258	.000[b]
	Residual	17929.187	225	79.685		
	Total	25342.530	229			

a. Dependent Variable: sleepy & assoc sensations scale

b. Predictors: (Constant), physical fitness, age, sex, HADS Depression

Evaluation of the Model Summary table:

The R square tells you how much of the variant in the dependent variable. In this case, the value r=.541 (54.1%) of the variance perceived sleep. This is a good result. The adjusted R Square value statistics corrects this value to provide a better estimate of the true population value. The Adjusted R Square on this data is.280. To assess the statistical significance we need to check the ANOVA to test the null hypothesis that multiple R in the population equals to 0. The data shows a Sig = .000 which is statistically significant

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	32.211	3.481		9.255	.000

	sex	-3.323	1.193	-.157	-2.786	.006
	age	-.121	.047	-.146	-2.604	.010
	HADS Depression	1.522	.208	.433	7.329	.000
	physical fitness	-.731	.364	-.119	-2.008	.046

Coefficients[a]

Model		95.0% Confidence Interval for B		Correlations		
		Lower Bound	Upper Bound	Zero-order	Partial	Part
1	(Constant)	25.352	39.069			
	sex	-5.673	-.973	-.199	-.183	-.156
	age	-.213	-.029	-.141	-.171	-.146
	HADS Depression	1.113	1.932	.482	.439	.411
	physical fitness	-1.447	-.014	-.267	-.133	-.113

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	sex	.986	1.014
	age	.998	1.002
	HADS Depression	.900	1.111
	physical fitness	.892	1.121

a. Dependent Variable: sleepy & assoc sensations scale

The Coefficient Table contributes to the prediction of the dependent variable. We look at the column labeled Beta Standardized Coefficient and looked for the largest beta coefficient. In this

case, the largest beta coefficient is .433 which is HADS depression with a Sig=.000 which is statistically significant.

Overall these results, answer the 2 questions we initially posted.Q1 – How well do variables: age, gender, fitrate and depress predict the level of sleepiness and associated symptoms? Q2 – Which variable best predict totsas? The statistical analysis indicates that all dependent variables provide a moderate to the high prediction of variance in the dependent variable. The HADS Depression with a Sig of .000 and ANOVA and R Square Sig = .000 correlates between the dependent and independent variable in predicting totsas. The HADS Depression showed a relationship with the totsas with the beta coefficient of .433. Since this result is above .3, which is preferred according to Pallant (2016, page 159) there will be no changes or removal of any of the variables. See table Variables entered/removed.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	physical fitness, age, sex, HADS Depression[b]	.	Enter

a. Dependent Variable: sleepy & assoc sensations scale

b. All requested variables entered.

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	sex	age	HADS Depression	ph fitness
1	1	4.027	1.000	.00	.02	.00	.02	
	2	.539	2.734	.00	.69	.00	.21	
	3	.343	3.424	.00	.28	.02	.58	

	4	.071	7.509	.00	.01	.62	.03
	5	.020	14.235	.99	.00	.35	.16

a. Dependent Variable: sleepy & assoc sensations scale

2. Run a hierarchical multiple regression procedure using the same variables you used in regression number one. Control for gender and age then examine the effects of physical fitness and depression on the level of sleepiness and associated symptoms.

a. State your research question.

If we control for the possible effect of age and gender (independent variable), is our set of the other variables HADs Depress and Fitrate (independent variable) still able to predict a significant amount of variance in totsas (dependent variable – sleep and associated symptoms)?

b. Present and interpret your findings. Copy and paste relevant computer output that displays results and provide a narrative summary of the output. Be sure to answer your research question.

To answer this research question, I will be using hierarchical multiple regression. Variables will be entered in steps of blocks. In the first block, I will force age and gender (sex). This has the effect of statistically controlling for these variables. Next, we enter the other independent variable into the model as a block (similar to the first block)

Below is my run…

```
GET
  FILE='C:\Users\rvaldres\Desktop\Valdres.sleep5ED.new.sav..sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING PAIRWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT totsas
  /METHOD=ENTER sex age
  /METHOD=ENTER depress fitrate
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS NORMPROB(ZRESID)
  /SAVE MAHAL COOK.


[DataSet1] C:\Users\rvaldres\Desktop\Valdres.sleep5ED.new.sav..sav
```

Descriptive Statistics

	Mean	Std. Deviation	N
sleepy & assoc sensations scale	26.0451	10.52051	2
sex	.45	.498	2
age	43.87	12.684	2
HADS Depression	3.50	2.993	2
physical fitness	6.42	1.717	2

Correlations

		sleepy & assoc sensations scale	sex	age	HADS Depression
Pearson Correlation	sleepy & assoc sensations scale	1.000	-.199	-.141	.482
	sex	-.199	1.000	-.017	-.071
	age	-.141	-.017	1.000	-.004
	HADS Depression	.482	-.071	-.004	1.000
	physical fitness	-.267	.110	-.039	-.314
Sig. (1-tailed)	sleepy & assoc sensations scale	.	.001	.017	.000
	sex	.001	.	.393	.124
	age	.017	.393	.	.473
	HADS Depression	.000	.124	.473	.
	physical fitness	.000	.001	.000	.000
N	sleepy & assoc sensations scale	2	2	2	2
	sex	2	2	2	2
	age	2	2	2	2
	HADS Depression	2	2	2	2
	physical fitness	2	2	2	2

	physical fitness	.000	.037	.271	.000
N	sleepy & assoc sensations scale	251	251	230	249
	sex	251	271	248	269
	age	230	248	248	246
	HADS Depression	249	269	246	269
	physical fitness	247	266	243	265

Correlations

		physical fitness
Pearson Correlation	sleepy & assoc sensations scale	-.267
	sex	.110
	age	-.039
	HADS Depression	-.314
	physical fitness	1.000
Sig. (1-tailed)	sleepy & assoc sensations scale	.000
	sex	.037
	age	.271
	HADS Depression	.000
	physical fitness	.
N	sleepy & assoc sensations scale	247
	sex	266
	age	243
	HADS Depression	265
	physical fitness	266

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	age, sex[b]	.	Enter

2	HADS Depression, physical fitness[b]	.	Enter

a. Dependent Variable: sleepy & assoc sensations scale

b. All requested variables entered.

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics		
					R Square Change	F Change	df1
1	.245[a]	.060	.052	10.243	.060	7.267	2
2	.541[b]	.293	.280	8.927	.232	36.948	2

Model Summary[c]

Model	Change Statistics	
	df2	Sig. F Change
1	227	.001
2	225	.000

a. Predictors: (Constant), age, sex

b. Predictors: (Constant), age, sex, HADS Depression, physical fitness

Dependent Variable: sleepy & assoc sensations scale

Evaluating the Model

Model 1 refers to the first block of independent variables (age and sex – control) and model 2 the second set of block. Check the R Square values in the first model summary box. After the variables in the first block are entered, the overall model explains 6.0% of the variance (.60x100) (this is the R Square in the box). After block 2 variables (HADS Depression & Physical Fitness), the model as a whole explains 29.3% (.293x100). The second R Square value includes all the variables from both blocks, not just those included in the second step.

Next, to find out how much of this overall variance is explained by our variables of interest (totsas) after the effects of age and sex are removed, I need to look at **R Square Change**. The output I have shown 60% (.60 X 100). This means that the totsas explain an additional 60% of the variance, even when the effects of age and sex are statistically controlled. This is statistically significant as indicated in **Sig F Change** with the value of .001

ANOVA table indicates that the model as a whole is significant = .001

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1524.909	2	762.455	7.267	.001[b]
	Residual	23817.621	227	104.923		
	Total	25342.530	229			
2	Regression	7413.343	4	1853.336	23.258	.000[c]
	Residual	17929.187	225	79.685		
	Total	25342.530	229			

a. Dependent Variable: sleepy & assoc sensations scale

b. Predictors: (Constant), age, sex

c. Predictors: (Constant), age, sex, HADS Depression, physical fitness

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	33.184	2.521		13.162	.000
	sex	-4.246	1.359	-.201	-3.123	.002
	age	-.120	.053	-.144	-2.240	.026
2	(Constant)	32.211	3.481		9.255	.000
	sex	-3.323	1.193	-.157	-2.786	.006
	age	-.121	.047	-.146	-2.604	.010
	HADS Depression	1.522	.208	.433	7.329	.000
	physical fitness	-.731	.364	-.119	-2.008	.046

Coefficients[a]

Model		Correlations			Collinearity Statistics	
		Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)					
	sex	-.199	-.203	-.201	1.000	1.000
	age	-.141	-.147	-.144	1.000	1.000
2	(Constant)					
	sex	-.199	-.183	-.156	.986	1.014

age	-.141	-.171	-.146	.998	1.002
HADS Depression	.482	.439	.411	.900	1.111
physical fitness	-.267	-.133	-.113	.892	1.121

a. Dependent Variable: sleepy & assoc sensations scale

Evaluating the independent variables:

To find out how well each of the variables contributes to the final equation, we need to look at the **Coefficient Table** in the model 2 row

This summarizes the results with all the variables entered into the equation. Looking through the **Sig** column, there are only two variables that make a unique statistically significant

Contribution (less than .05). In order of importance (based on their beta values), they are:

HADS Depression beta =.433, physical fitness beta= .119

The HADS depression has a more significant effect on totsas compared to physical fitness. These beta values represent the uniques contribution of each variable when the

Overlapping effects of all other variables are statistically removed (Pallant, 2016, page 167)

Overall summary:

Hierarchical multiple regression statistical analysis was used to assess the ability of two control measures (HADS Depression and physical fitness) to predict

Sleepiness and its associated symptoms (totsas), after controlling for the influence of age and gender/sex. Preliminary analyses were conducted to ensure no violation of the

Assumptions of normality, linearity, multicollinearity, and homoscedasticity. Age and sex were entered in step 1 with the R-square of 0.60 (model 1 table). This answers

My research question. The analysis shows HADS depression made a significant impact than the physical fitness in predicting the amount of variance in totsas. The control (age and sex)

Contributed to the significant result of this analysis. In conclusion, my research question: " If we control for the possible effects of age and sex, is our set of variables (HAD depression and physical fitness still able to predict the significant amount of variance in totsas? " was answered.

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics: Tolerance	Collinearity Statistics: VIF
1	HADS Depression	.470[b]	8.303	.000	.483	.995	1.005
	physical fitness	-.254[b]	-4.046	.000	-.260	.987	1.014

Excluded Variables[a]

Model		Collinearity Statistics: Minimum Tolerance
1	HADS Depression	.995
	physical fitness	.987

a. Dependent Variable: sleepy & assoc sensations scale

b. Predictors in the Model: (Constant), age, sex

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions: (Constant)	sex	age	HADS Depression
1	1	2.522	1.000	.01	.06	.01	
	2	.440	2.395	.02	.92	.03	
	3	.038	8.112	.97	.02	.96	
2	1	4.027	1.000	.00	.02	.00	.02
	2	.539	2.734	.00	.69	.00	.21
	3	.343	3.424	.00	.28	.02	.58
	4	.071	7.509	.00	.01	.62	.03

5	.020	14.235	.99	.00	.35	.16

Collinearity Diagnostics

Model	Dimension	Variance Proportions: physical fitness
1	1	
	2	
	3	
2	1	.00
	2	.00
	3	.03
	4	.34
	5	.63

a. Dependent Variable: sleepy & assoc sensations scale

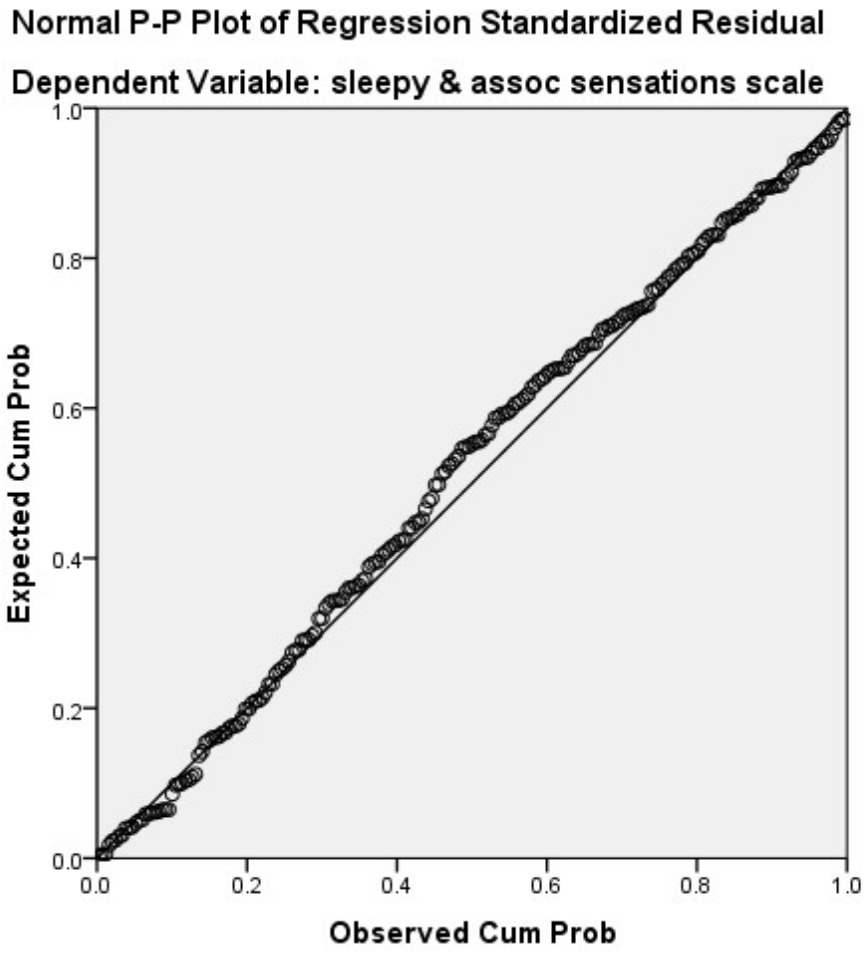
Residuals Statistics[a]

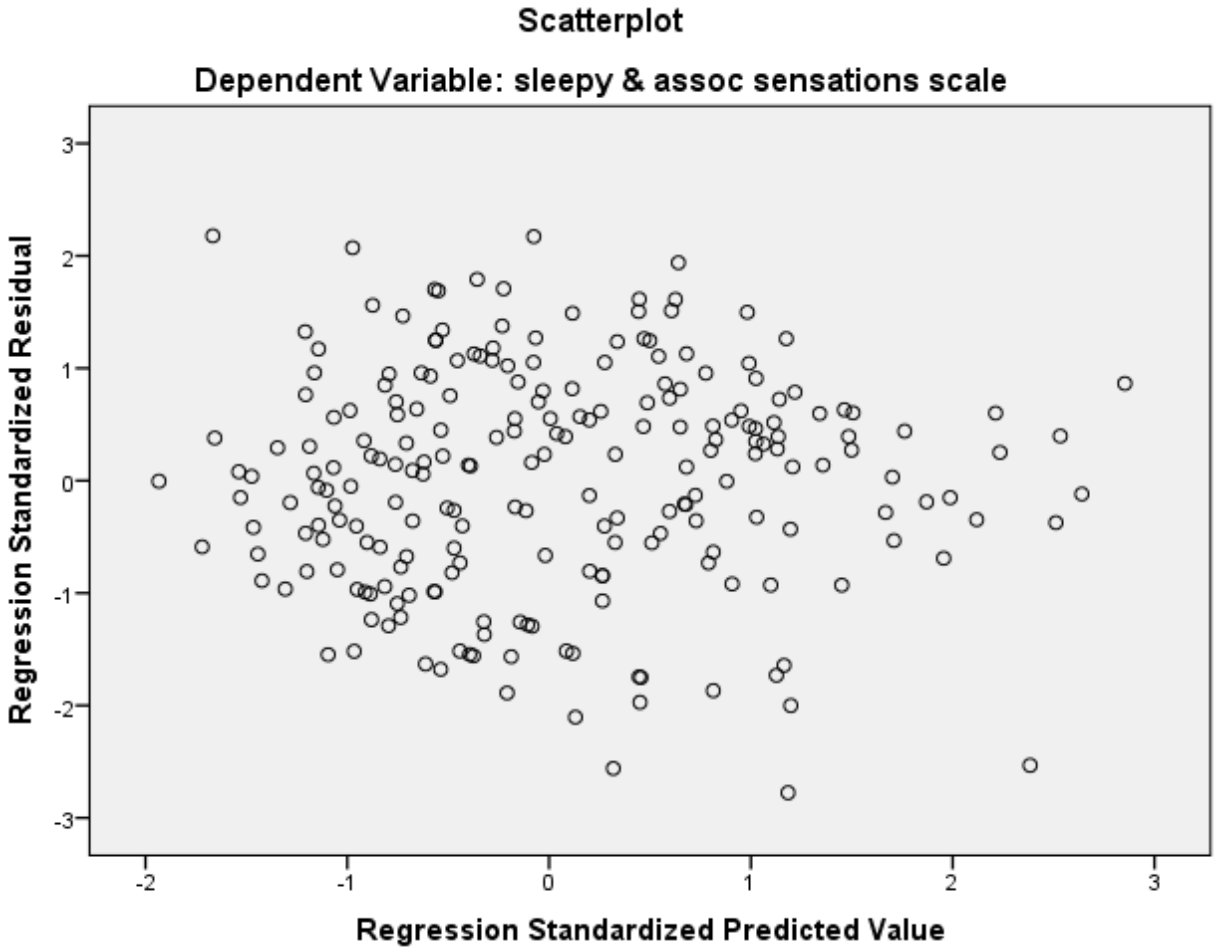
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	15.04	42.28	26.15	5.653	242
Std. Predicted Value	-1.933	2.853	.018	.993	242
Standard Error of Predicted Value	.816	2.269	1.281	.283	242
Adjusted Predicted Value	15.04	41.95	26.27	5.577	225
Residual	-24.781	19.438	.178	8.839	225
Std. Residual	-2.776	2.178	.020	.990	225
Stud. Residual	-2.810	2.207	.020	1.002	225
Deleted Residual	-25.393	19.980	.172	9.050	225

Stud. Deleted Residual	-2.854	2.226	.019	1.006	225
Mahal. Distance	.919	13.804	3.949	2.314	242
Cook's Distance	.000	.075	.005	.009	225
Centered Leverage Value	.004	.060	.017	.010	242

a. Dependent Variable: sleepy & assoc sensations scale

Charts




Scatterplot
Dependent Variable: sleepy & assoc sensations scale
Regression Standardized Residual
Regression Standardized Predicted Value
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3

References

Pallant, J. (2016). *SPSS Survival Manual* (6th ed.). NY: McGraw Hill Education.