Analysis on the Relationship Between Perceived Mental Health and Duration of Sleep in Young Canadians

Bojing Yao
bojing.yao@torontomu.ca,
Department of Computer Science
Toronto Metropolitan University
Toronto, Canada

***Abstract* - This project explores the relationship between sleep duration and perceived mental health among young Canadians through the use of the 2017-2018 Canadian Community Health Survey. With a rising trend in mental health issues within Canada, this study aims to understand if longer sleep hours correlate with a better perceived mental health outcomes. Through multivariate analyses that incorporates variables such as perceived stress, income, and education levels, this project has identified significant association between sleep duration, stress, income, education, and mental health. Key findings suggest that sleep within the recommended 7-9 hours, a reduced stress levels, higher income, and higher education attainment are linked to improved mental health perceptions. This study suggests improvement to mental well-being by improving the identified variables.**

I. Introduction

There is a growing concern over mental health in Canada, particularly in the context of young Canadians. Smetanin (2011) illustrates a rapidly escalating prevalence of mental disorders in Canada, underscoring a decade-long rise of mental disorders that estimates 1 in 5 Canadians to be affected by mental disorders. The rapidly decreasing mental health status requires the potential causes to be examined deeper, and a main factor could be the hours of sleep. Therefore, this project aims to discover any correlation between the amount of sleep and individual's perceived mental health status for young Canadians between the age of 18 to 24.

A literature review has been conducted based on researches that uses similar variables from the Canadian Community Health Survey (CCHS) dataset. Robert & Gilkinson (2012) looked into how income and perceptions of the settlement process are correlated to having a better mental health status in immigrants using a logistic regression analysis. Similarly, Geda & Feng (2022) studied the relationship between mental health, stress, and alcohol consumption, through a binomial regression model to correlate these relationships. Wang (2022) performed a study on the variations in sleep duration between the population and the recommended sleep duration guideline of 7-9 hours. In particular, Robert & Gilkinson (2012) has obtained a strong correlation while restructing the perceived mental health status variable to be re-categorized to be excellent, very good or fair versus others.

This literature review confirms the feasibility of this study, and demonstrates how the variables can be re-organized to produce a clearer result. Moreover, due to the categorical nature of the dataset, and the variables selected for this particular project, the literatures has shown the analysis model that is suitable for this project to be logistic regression model.

II. Methods

The CCHS dataset is a cross-sectional dataset for a variety of participants. In order to perform an analysis identifying the relationship between any two variables, rigorous data filtering must be performed in order for the data to be relevant and contains no outliers. A confounding analysis must also be performed, and the co-variates identified. A confounder variable is one that is associated with both the independent variable and the outcome, which in this case may be variables that can affect both the sleep duration in young adults and the perceived mental health status in young adults.

Confounders in this project are identified through logical analysis and verified using logistic regression to confirm it's relationship to the variables. The confounders are controlled through including them as covariates in the regression model. This project performs multivariate analysis, including multiple covariates to better understand if the relationship between the independent variable and the outcome is a composite one. The various covariates are resulted from literature review and practical considerations. There are also universal covariates, such as education level, food security and income level.

The outcome variable in this study is the perceived mental health status (GEN_015), and the independent variables are the duration of sleep (SLPG005), the perceived life stress (GEN_020), the income level (INCDGPER), the education level (EHG2DVR3), the food security (FSCDVAFS) and the perceived health (GEN_005). The perceived health and the perceived life stress is selected to be the confounder to be analyzed and included in the model as covariates. GEN_020 and GEN_015 and GEN_005 are transformed to "high risk" and "low risk" to better demonstrate any possible correlation between the independent variable and the outcome variable. The size of the filtered dataset cannot support the 11 categories for sleep duration, therefore, the SLPG005 variable is reduced to 3 categories based on the recommended sleep duration of 7-9 hours. The re-structured SLPG005 is consists of "below recommended duration", "within recommended duration" and "more than recommended duration".

The dataset is filtered based on the variable of people who participated in the survey (DOSLP) and people that are within the age 18-24 (DHHGAGE). There are also a set of exclusion categories for each individual variable of interest to this project, which also serves as the exclusion criteria for the dataset. The filter has shown there are 3382 participants suitable for this project, which is 3.0% of the entire dataset. A graph of the filtered data is presented.

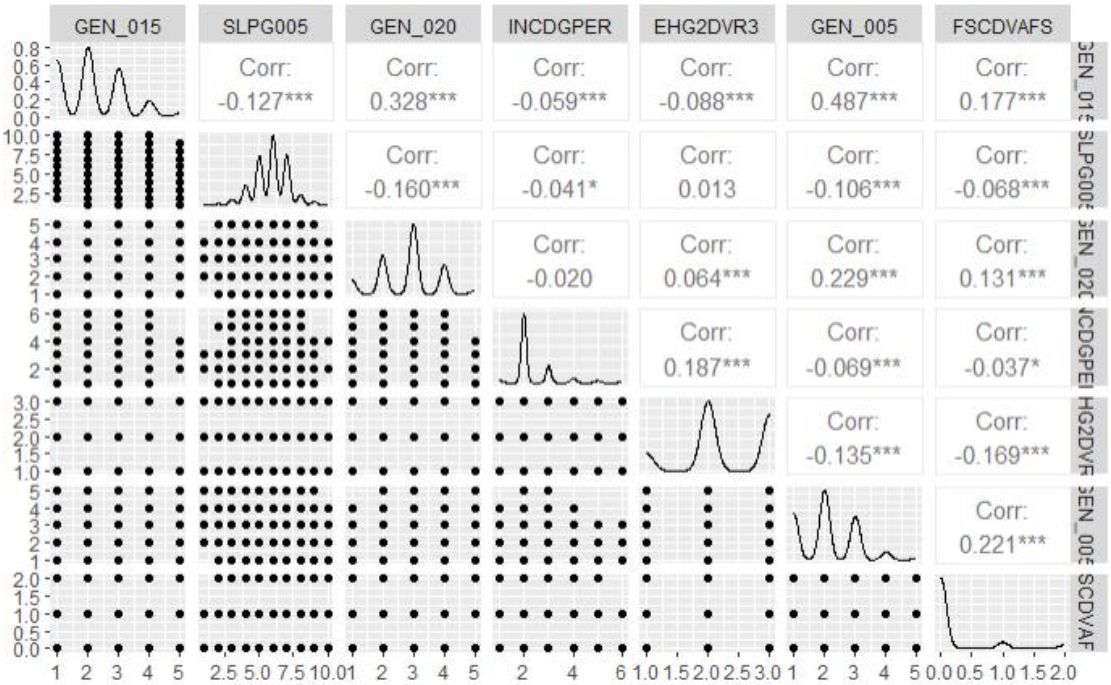

Figure 1: Dataset overview with pairs analysis between all variables

A neural networks model will be used to be trained on the subset of data suitable for this project, in order to evaluate if the covariates can be used to predict if an individual's perceived mental health status.

III. Analysis

The analysis result for this project is composed of the initial Chi-squared test between the various covariates and the outcome variable. The Chi-squared test is used to identify correlations between the covariates and the outcome variable due to the outcome variable being binary and categorical, while the covariates are categorical. The X-squared values suggests that the covariates selected in this project all have association to the outcome variable, while the p-value suggests that the null hypothesis can be rejected and the association is realistic. Out of all covariates, the health level has an overwhelmingly strong association, with a X-squared value of over 3000. The second strongly associated variable is the perceived stress in life, being higher than the rest of the variables.

Despite having an overwhelmingly strong association, the perceived health level is not included in the binomial regression model due to it being too broad, and the variable may be a mediator, having a direct impact on hours of sleep or perceived mental health, due to people with worse health generally suffers from a worse mental conditions and having poor sleep.

GEN_015	x-squared	p-value
SLPG005 - Duration of Sleep	168.45	2.2e-16
GEN_020 - Perceived life stress	659.11	2.2e-16
INCDGPER - Income interval	177.95	2.2e-16
EHG2DVR3 - Highest education level achieved	115.52	1.216e-11
GEN_00 - Perceived health	3417	2.2e-16
FSCDVAFS - Food security	83.644	2.2e-16

Table 1 - Initial Analysis of Significance

A binomial logistic regression model is used with SLPG005, GEN_020, INCDGPER, EHG2DVR3 and FSCDVAFS as the input with GEN_015 as the output. The output of the model is presented in Table 2. The resulting odds ratio incorporates the error into it and produces a 95% confidence interval range. The significance of each category is indicated using the asterisk mark, with *** being the most significant category. The resulting p-value for the majority of the categories are fairly large, which does not indicate that the null hypothesis can be excluded from the relationship. The most impactful variables as indicated from the model seems to be the 7-9 hours category in the sleep duration variable, the high category in perceived life stress, all categories from education levels and the second highest income level, and all categories from the food security variable. The majority of the categories from the income level variable has a fairly high error range and a high p-value, suggesting the correlation not being significant.

Perceived Mental Health		95% CI Odds		coefficient	Std.err	p-value
		from	to			
Sleep length	**< 7 hours (base)**	1				
	7-9 hours ***	0.633	0.872	-0.296	0.081	0.000283
	> 9 hours	0.591	1.101	-0.213	0.158	0.176482
Perceived life stress	**Low (base)**	**1**				
	High ***	1.484	2.126	0.574	0.091	3.55e-10
Income level	**No income (base)**	**1**				
	Less than $20k	0.662	1.624	0.036	0.228	0.872812
	$20K to $39k	0.704	1.826	0.126	0.243	0.603618
	$40k to $59k	0.568	1.791	0.008	0.292	0.976176
	$60k to $79k	0.205	1.044	-0.769	0.414	0.063566
	$80k or more	0.207	1.349	-0.636	0.477	0.182712
Education	**Less than secondary school (base)**	**1**				
	Secondary school **	0.416	0.658	-0.646	0.116	2.93e-08
	Post-secondary ***	0.338	0.548	-0.842	0.123	8.22e-12
Food Security	S**ecure (base)**	**1**				
	Moderately Insecure ***	1.593	2.664	0.722	0.131	3.59e-08
	Severely Insecure ***	2.055	4.046	1.059	0.172	8.90e-10

Table 2 - Logistic Regression Result

The confounder analysis in Table 3 and Table 4 analyzes the correlation between the confounder variable with both the outcome variable and the independent variable. The analysis is performed using binary logistic regression and multinomial logistic regression. Result shows that the presumed confounders have strong correlations to the outcome variable and the independent variable. The p-value for the binary logistic regression for both analysis indicates significance, while the decrease in coefficients as the category increases and the error also indicates some degree of correlation.

Outcome Variable	Model	Coefficient	Std. Error	z-value	p-value
Life stress (0)	Logistic Regression - Binary	1			
Life stress (1)	Logistic Regression - Binary	1.702	0.080	21.25	2e-16
Sleep duration (0)	Logistic Regression - Multinomial	1			
Sleep duration (1)	Logistic Regression - Multinomial	-0.433	0.078		
Sleep duration (2)	Logistic Regression - Multinomial	-0.169	0.151		

Table 3 - Confounding Analysis on GEN_005

Outcome Variable	Model	Coefficient	Std. Error	Z-value	P-value
Sleep duration (0)	Logistic Regression - Multinomial	1			
Sleep duration (1)	Logistic Regression - Multinomial	-0.729	0.088		
Sleep duration (2)	Logistic Regression - Multinomial	-0.648	0.183		
Life stress (0)	Logistic Regression - Binary	1			
Life stress (1)	Logistic Regression - Binary	0.990	0.086	11.49	2e-16

Table 4 - Confounding Analysis on GEN_020

To further test the theory of this project, a neural networks model is constructed with a summary as described in Figure 2. The variables with significant correlation to the outcome variable is taken as input to the neural network model. The model is used as a predictor with 2 hidden layers , an input layer and an output layer with one neuron. The training reaches the peak accuracy at 5 epochs. The accuracy obtained on testing data is 100%.

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 64)	384
dense_4 (Dense)	(None, 32)	2080
dense_5 (Dense)	(None, 1)	33

Total params: 2497 (9.75 KB)
Trainable params: 2497 (9.75 KB)
Non-trainable params: 0 (0.00 Byte)

Figure 2 - Neural Networks model for verifying the hypothesis.

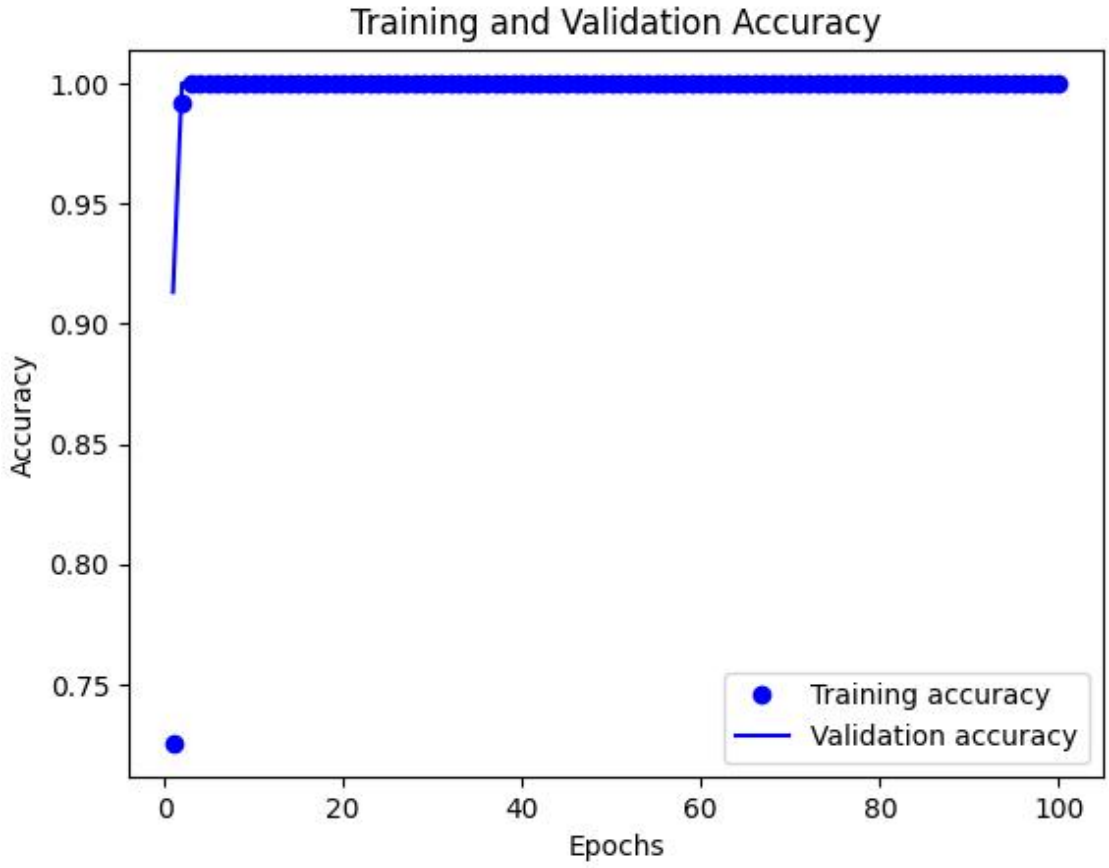

Figure 3 - Training and Validation Accuracy of Neural Networks model.

IV. Result

The results obtained shows a 95% confidence interval for the odds ratio of each category of having a "high risk" perceived mental health status, in comparison with the base category. The result shows that having a sleep duration within the recommended 7-9 hours reduces the likelihood of having a bad perceived mental health status by around 10%-20%. The perceived life stress is strongly correlated to perceived mental health, suggesting that it is necessary for this confounder to be controlled and included in the model. The result indicates that it is 1.5 to 2 times as likely for an individual with high life stress to have a high risk perceived mental health status, in comparison with individual who has low life stress. The results for the income levels are not significant as the p-value are too high in general to reject the null hypothesis, however, it would be accurate to say that people with an income of $60k to $79k would generally be less likely to have a bad mental health status than people with no income. The education level is very strongly correlated to perceived mental health status, with having a higher level education correlating to less odds of having a "high risk" mental health status. Food security also strongly correlates to the perceived mental health status, with less food security correlating to having a significantly increase in likelihood of having a bad mental health status.

In general, the result obtained from the models suggest that the perceived mental health status is more correlated with the general life stress, the education levels, and food security than hours of sleep. However, sleeping within the recommended 7-9 hours still decreases the odds of having a bad perceived mental health status, according to the odds ratio within the 95% confidence interval. The result also indicates that the correlation between income level and perceived mental health status is not very significant, suggesting that the level of income is not as important as to change a Canadian's mental health states. An unexpected discovery is that having a higher education level and having a secure food source is very important to secure a good mental health states, as having a university level education is 33% to 55% as likely as someone with less than a secondary school degree to have a bad perceived mental health status, and being food insecure results in a 2-4 times likelihood for a person to have a bad perceived mental health status. The neural network model confirms the hypothesis of this project, that an individual's perceived mental health status can be determined by the combination of all the covariates introduced in this project. The neural network model can be used to predict an individual's mental health status and can be used to identify high risk individuals and provide additional help to them.

The results obtained from this project can be used to determine a reasonable threshold to control the mental disorder issues that the Canadian public faces. This analysis shows that a decrease in mental health status is a combined effort of not having a high level education degree, not being food secure, not getting the recommended hours of sleep, and having a high stress in life. Given this result, the Canadian government can attempt to improve the overall perceived mental health status of the public through providing a higher minimum wage and reduce the cost of attending an university or college level program. The government can also focus on providing a higher level of food security.

A main limitation that this project faced is the categorization. The resulting dataset after the extensive filtering on suitable candidates from the original CCHS dataset is less than expected, and the levels of income variable suffered from a lack of data for some of it's categories. This may result in the data for the level of income to be less accurately reflective.

References

Geda, N. & Feng, C. (2022). Perceived mental health, work, and life stress in association with the amount of weekly alcohol consumption among Canadian adults who have ever drank. *BMC Public Health.* DOI: https://doi.org/10.1186/s12889-022-14240-3

Robert & Gilkinson (2012). Mental health and well-being of recent immigrants in Canada: Evidence from the Longitudinal Survey of Immigrants to Canada. *Research and Evaluation*. https://www.canada.ca/content/dam/ircc/migration/ircc/english/pdf/research-stats/mental-health.pdf

Wang, C. et al (2022). Sleep behaviours among Canadian adults: Findings from the 2020 Canadian Community Health Survey healthy living rapid response module. Health Reports. https://www150.statcan.gc.ca/n1/en/pub/82-003-x/2022003/article/00001-eng.pdf?st=mHcmgPn0

Smetanin et al. (2011). The life and economic impact of major mental illnesses in Canada: 2011-2041. *Mental Health Commission of Canada. https://www.mentalhealthcommission.ca/wp-content/uploads/drupal/MHCC_Report_Base_Case_FINAL_ENG_0_0.pdf*