INTRODUCTION

By the early 1980s, scientific research clearly showed that personal health behaviors played a major role in premature morbidity and mortality. Initially developed in the early 1980s by the Centers for Disease Control and Prevention (CDC), the Behavioral Risk Factor Surveillance System (BRFSS) is a United States health survey that looks at behavioral risk factors associated with the leading causes of premature mortality and morbidity among adults. In this data analysis, we are considering the following variables from the BRFSS: 1) sex of the respondent; 2) age of the respondent; 3) height of the respondent without shoes; 4) weight of the respondent without shoes; 5) body mass index (BMI); 6) annual household income; 7) whether the respondent have had either a seasonal flu shot or a seasonal flu vaccine; 8) marital status; 9) whether the respondent have any disabilities (physical, mental or emotional); 10) average of hours of sleep, and; 11) whether the respondent snores.

This particular analysis is focused on the behavioral risk factors associated with weight of the individuals. A preliminary analysis of the data shows that both female individuals and male individual showed a similar range of values for the body mass index (Figure 1).

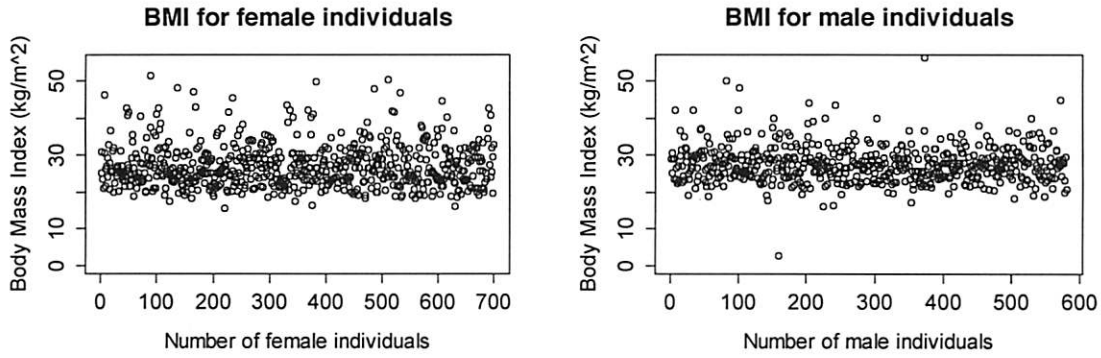

Figure 1. Body mass index for male and female individuals

This dataset has a sample of 701 female individuals from which: 564 are married and 137 never married. For the males, this dataset has a sample of 582 individuals from which 433 are married and 149 never married. The mean values for the body mass index of never married female and male individuals are 26.86 kg/m² and 26.03 kg/m², respectively; whereas the mean values for the body mass index of married female and male individuals are 27.19 kg/m² and 28.13 kg/m², respectively. Even though they might help us to guide our analysis, these sample averages values are not sufficient to statistically address the following questions of interest:

1. ***"Do married women tend to be heavier than never married women?"***
2. ***"Do married men tend to be heavier than never married men?"***

Additionally, based on medical evidence that suggests that lack of sleep may cause you to gain weight, a third question for this analysis could be:

3. ***"Do people who sleep less than 7 hours tend to be heavier than people who sleep more than 7 hours?"***

To be able to answer these questions we need to use a set of statistical tools that is described on the following section.

METHODS

In order to perform a two sample t-test – and obtain a valid result – we have to check some mathematical assumptions of the two sample t-test. There are three assumptions: Independence, Normality and Equal Standard Deviations. In practice this assumptions don't have to be met, since the t-tools are robust when some of these assumptions are violated.

Question of interest 1 and 2

Independence

The independence assumption is met for both male and female samples, since the procedure adopted by the BRFSS for selecting the sample is based in the random sampling model.

Equal Standard Deviations

We check this assumption by looking at the histograms of samples (Figure 2):

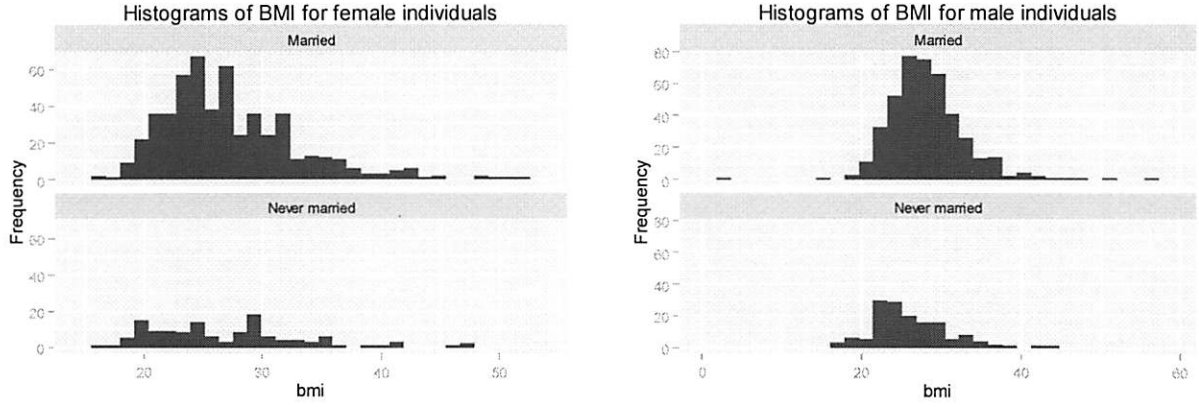

Figure 2. Histograms of samples

Looking at the histograms of the BMI for married and never married males and females suggests a somewhat similar variance. Additionally, the standard deviation ratio for female individuals is 1.12 and for males is 1.02 (with 1 being equal standard deviations). Hence, the assumption of equal standard deviations is satisfied (even though the sample sizes are different). However, examining the histogram of BMI for males reveals some outliers: one with very low body mass index and one with very large body mass index. The t-tools are not resistant to outliers and these values may be influential in the analysis.

Normal Probability Plot

The normal probability plots for married and never married male and female individuals can be seen in Figure 3.

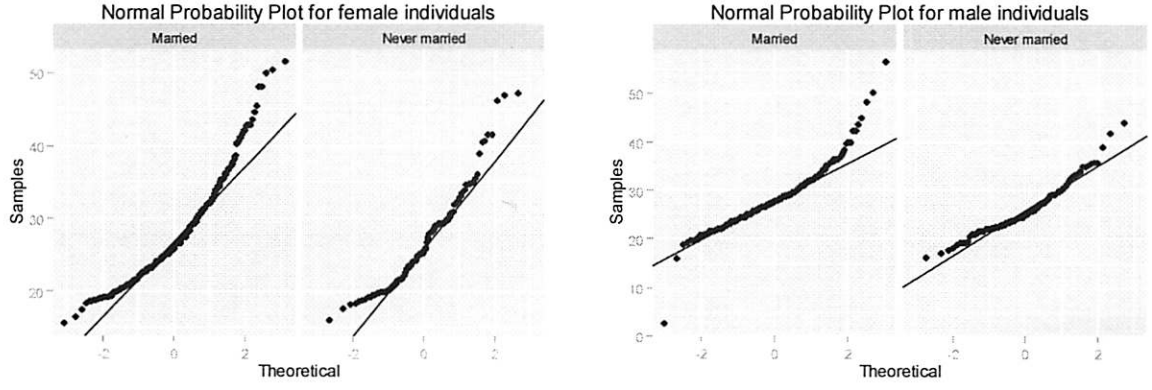

Figure 3. Normal Probability Plots

The normal probability plots suggest that these sample data aren't exactly normal. However, the condition of normality is robust due to the central limit theorem and there is a relatively large sample size for both male and female individuals. Thus, we should not worry about this particular condition.

Question of interest 3

Regarding the average hours of sleep, the histograms suggests a somewhat similar variance (Figure 4). Additionally, the standard deviation ratio for people who sleep less than 7 hours and for people who sleep more than 7 hours is 1.30 (with 1 being equal standard deviations). Hence, the assumption of equal standard deviations is satisfied. Similarly to the previous analysis, the normal probability plots suggest that these sample data aren't exactly normal. However, the condition of normality is robust due to the central limit theorem and there is also a relatively large sample size for people who sleep more than 7 hours (722) and less than 7 hours (685). We can also see that the sample sizes are roughly equal in this case. Hence, the two sample t-test will be fairly valid.

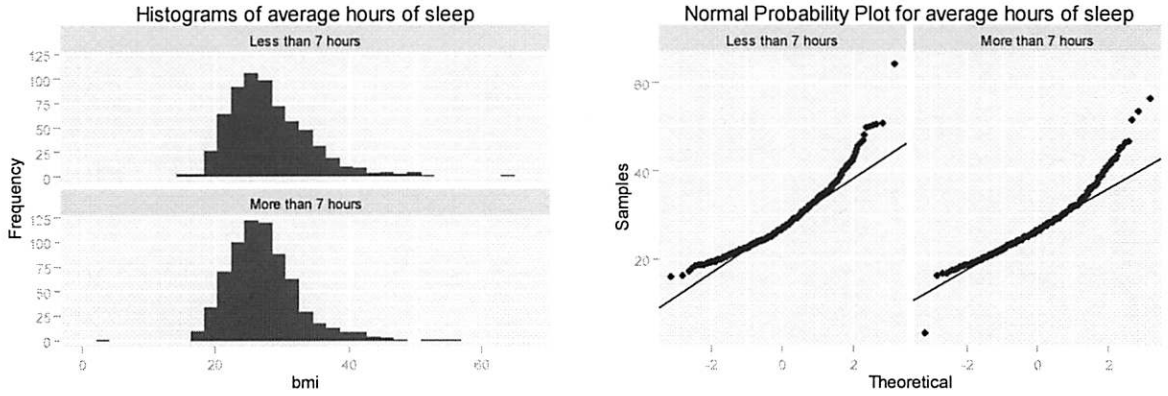

Figure 4. Histograms of samples and Normal Probability Plots for people who sleep more and less than 7 hours.

Sometimes the assumptions can be met by transforming the data. A particularly useful transformation is the logarithmic transformation. However, in this case, the data was not transformed since I am assuming that the assumptions were met for both cases (Questions 1 and 2 and Question 3). Hence, I am performing a **two-sample t-test** to be able to answer the questions of interest.

The Table 1 shows a summary of the two sample t-test performed for each question of interest.

Table1. Two sample t-test results for the questions of interest

Question	Test statistic	Degrees of freedom	p-value	95% Confidence interval		Sample estimates per group	
				Upper limit	Lower Limit		
1	0.5684	655	0.57	1.46	-0.80	27.18 kg/m² *	26.86 kg/m² **
2	4.5123	571	< 0.0001	3.00	1.18	28.13 kg/m² *	26.03 kg/m² **
3	2.9204	1340	0.0035	1.52	0.30	28.05 kg/m² [a]	27.14 kg/m² [b]

Groups: * Married ** Never married [a] Less than 7 hours [b] More than 7 hours

A preliminary analysis of the Table 1 shows that the p-value for the married and never married females is considerably large compared to the others. This value represents the probability of seeing a test statistic as or more extreme than the observed assuming that the null hypothesis is true (population means are the same). So, in this particular case, the p-value is larger than the significance level of 0.05. This means that we failed to reject the null hypothesis. A more detailed statistical summary for each question of interest is provided in the following section.

STATISTICAL SUMMARY

The results of the two-sample t-test (Table 1) can be used to answer the questions of interest.

Do married women tend to be heavier than never married women?

There is **no evidence** that the mean body mass index for married women **is not equal** to the mean body mass of never married women (two sample t-test, two sided p-value = 0.57). In other words, we failed to reject the null hypothesis (null: the means are equal). Thus, married women do not tend to be heavier than never married women.

The mean body mass index of married women is estimated to be **0.32 kg/m²** higher than the mean body mass index of never married women.

With 95% confidence the mean body mass index for the population of married women is between **-0.80** and **1.46 kg/m²** higher than the mean body mass index for the population of never married women.

Do married men tend to be heavier than never married men?

In this case, there is **convincing evidence** that mean body mass index for married men **is not equal** to the mean body mass of never married men (two sample t-test, two sided p-value < 0.0001). So, we reject the null hypothesis.

The mean body mass index of married men is estimated to be **2.1 kg/m²** higher than the mean body mass index of never married men. So, we conclude that married men tend to be heavier than never married men.

With 95% confidence the mean body mass index for the population of married men is between **1.18** and **3.0 kg/m²** higher than the mean body mass index for the population of never married men.

Do people who sleep less than 7 hours tend to be heavier than people who sleep more than 7 hours?

For this third situation, there **is convincing evidence** that the mean body mass index for people who sleep more than the average of 7 hours **is not equal** to the mean body mass index for people who sleep less than the average of 7 hours (two sample t-test, two sided p-value = 0.0035). So, we reject the hypothesis that the means are equal for both populations.

The mean body mass index for people who sleep more than the average of 7 hours is estimated to be **0.91 kg/m²** smaller than the mean body mass index for people who sleep less than the average of 7 hours. Hence, we conclude that people who sleep less than the average of 7 hours tend to be heavier than people who sleep more than the average of 7 hours.

With 95% confidence the mean body mass index for the population of people who sleep less than the average of 7 hours is between **0.30** and **1.52 kg/m²** higher than the mean body mass index for the population of people who sleep more than the average of 7 hours.