# SAS Assignment 3

## Data Analysis using SAS & Sample Weights

Using your data file from Assignment 2, report the following:

**Part 1:**

**Among participants born in 1959 or earlier:**

1. Report the descriptive statistics of participants in your sample

Age (M, SD), Gender, Race, Hispanic, Education, Coupleness

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analysis Variable : NAGE AGE AT 2012 INTERVIEW | | | | |
| N | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 19327 | 68.0311999 | 10.7913427 | 52.0000000 | 103.0000000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender Recode | | | | |
| female | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Male | 8839 | 42.79 | 8839 | 42.79 |
| Female | 11820 | 57.21 | 20659 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RACE/ETHNICITY | | | | |
| RACE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| White | 15090 | 73.22 | 15090 | 73.22 |
| Black | 3938 | 19.11 | 19028 | 92.33 |
| Other | 1580 | 7.67 | 20608 | 100.00 |
| Frequency Missing = 51 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HISPANICITY TYPE | | | | |
| HISPANIC | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Mexican | 1552 | 7.52 | 1552 | 7.52 |
| Other | 1034 | 5.01 | 2586 | 12.53 |
| Unknown | 35 | 0.17 | 2621 | 12.70 |
| Non-Hispanic | 18010 | 87.30 | 20631 | 100.00 |
| Frequency Missing = 28 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Degree combined; 1 = Associates or higher | | | | |
| DEGREE2 | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No College Degree | 14954 | 77.05 | 14954 | 77.05 |
| Obtained an associates degree or higher | 4454 | 22.95 | 19408 | 100.00 |
| Frequency Missing = 1251 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2012 WHETHER COUPLED OR PARTNERED | | | | |
| NCOUPLE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No | 8184 | 39.61 | 8184 | 39.61 |
| Yes | 12475 | 60.39 | 20659 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2014 WHETHER COUPLED OR PARTNERED | | | | |
| OCOUPLE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No | 8082 | 41.50 | 8082 | 41.50 |
| Yes | 11393 | 58.50 | 19475 | 100.00 |
| Frequency Missing = 1184 | | | | |

1. What is the mean and standard deviation for depressive symptoms in 2012 and 2014?

|  |  |  |
| --- | --- | --- |
| Label | Mean | Std Dev |
| |  | | --- | | **CESD-8 composite score 2012** | | **CESD-8 composite score 2014** | | |  | | --- | | 1.5330367 | | 1.4995325 | | |  | | --- | | 2.0450797 | | 2.0340994 | |

1. Is there a significant difference in depressive symptoms between men and women? Examine this for both the 2012 and 2014 waves.

**2012 Wave:**

Men mean depressive score = 1.3209

Women mean depressive score = 1.6874

Yes, t-value = -11.98, p<0.0001. Men experience statistically significantly less depression in the 2012 wave.

**2014 Wave:**

Men mean depressive score = 1.2812

Women mean depressive score = 1.6542

Yes, t-value = -11.49, p<0.0001. Men experience statistically significantly less depression in the 2014 wave.

1. Is there a significant difference in depressive symptoms between those who are working vs. those who are not?

**2012 Wave:**

Employed mean depressive score = 1.0545

Unemployed mean depressive score = 1.7450

Yes, t-value = 20.84, p<0.0001. Employed experience statistically significantly less depression in the 2012 wave.

**2014 Wave:**

Employed mean depressive score = 1.1162

Unemployed mean depressive score = 1.7874

Yes, t-value = 21.75, p<0.0001. Employed experience statistically significantly less depression in the 2012 wave.

1. What is the N and correlation between depressive symptoms in 2012 and self-rated health in 2014?

**2012 Wave**

N = 16231, Pearsons Correlation = -0.37698

What is the N and correlation between depressive symptoms in 2014 and self-rated health in 2014?

**2014 Wave**

N = 16,031, Pearsons Correlation = -0.428

1. How does the correlation change between 2012 and 2014, if at all? Why?

The correlation from 2012 to 2014 became more negative, indicating a stronger negative relationship between depressive symptoms in 2014 when correlated with self-rated health in 2014 as opposed its correlation in 2012. The main reason I can think of the difference in correlations between these two time points is because overall mean depression scores dropped in 2014, indicating individuals are less depressed, thus the trend was likely magnified and the negative correlation became stronger.

1. What is the correlation between depressive symptoms in 2012 and the psychosocial scale you created?

**Impulse Scale (Higher scores = more impulsivity):**

Pearson’s Correlation = -0.18167

1. What is the correlation between depressive symptoms in 2014 and the psychosocial scale you created?

**Impulse Scale (Higher scores = more impulsivity):**

Pearson’s Correlation = -0.15135

1. How do scores on the psychosocial scale you created relate to self-rated health in 2014?

When running a simple linear regression model, using Impulse as a predictor for self-rated health at 2014, there is a positive relationship. Using self-rated health as the outcome variable, impulsivity scores are positively related to self-rated health:Y(self-rated health) = 2.390 + 0.167(Ximpulse). Our model was overall predictive of self-rated health F(1, 6203) = 116.22, P<0.0001. In total, this model explains 1.8% of the variance in self-rated health. Additionally, our simple linear model indicates that impulsivity is a statistically significant predictor of self-rated health, t(6203) = 10.78, p<0.0001.

Therefore, for every 1 unit increase in the impulse score, we expect to see a 0.167 in someone’s self-rated health score. This indicates that individuals with higher impulse scores rate there health at higher levels.

1. Does the relation between psychosocial scores and self-rated health change when you control for demographic variables (e.g., age, race, gender, education, coupleness)?

Controlling for Age, race, whether coupled at 2014, gender, and degree, the relationship between Impulsivity & self-rated health remained positive and significant. Additionally, the overall multiple regression model, with the control variables remained significant in predicting self-rated health: F(6, 5758) = 86.20, p <0.0001. In total, this model explains 8.2% of the variance in self-rated health.

Additionally, when holding all control variables constant, Impulsivity remained a significant predictor of self- rated health with a beta of 0.15, t(5758) = 20.53, p < 0.0001. For every one unit increase in impulsivity, there is an expected 0.15 increase in someone’s health-rating.

**Part 2:** Repeat the analyses above using the 2012 sample weights (NWGTR). For most analyses, all you need to do is add the statement WEIGHT NWGTR; to your analysis. Answer the same questions now using the sample weights.

1. Report the descriptive statistics of participants in your sample

Age (M, SD), Gender, Race, Hispanic, Education, Coupleness

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analysis Variable : NAGE AGE AT 2012 INTERVIEW | | | | |
| N | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 19327 | 65.7183091 | 689.5661029 | 52.0000000 | 103.0000000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender Recode | | | | |
| female | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Male | 41680606 | 45.95 | 41680606 | 45.95 |
| Female | 49018154 | 54.05 | 90698760 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RACE/ETHNICITY | | | | |
| RACE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| White | 75281931 | 83.13 | 75281931 | 83.13 |
| Black | 9254875 | 10.22 | 84536806 | 93.35 |
| Other | 6019746 | 6.65 | 90556552 | 100.00 |
| Frequency Missing = 142208 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HISPANICITY TYPE | | | | |
| HISPANIC | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Mexican | 4888149 | 5.39 | 4888149 | 5.39 |
| Other | 2824334 | 3.12 | 7712483 | 8.51 |
| Unknown | 82681 | 0.09 | 7795164 | 8.60 |
| Non-Hispanic | 82813225 | 91.40 | 90608389 | 100.00 |
| Frequency Missing = 90371 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Degree combined; 1 = Associates or higher | | | | |
| DEGREE2 | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No Degree | 58922368 | 69.87 | 58922368 | 69.87 |
| Associates Degree or Higher | 25405982 | 30.13 | 84328350 | 100.00 |
| Frequency Missing = 6370410 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2012 WHETHER COUPLED OR PARTNERED | | | | |
| NCOUPLE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No | 31748546 | 35.00 | 31748546 | 35.00 |
| Yes | 58950214 | 65.00 | 90698760 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2014 WHETHER COUPLED OR PARTNERED | | | | |
| OCOUPLE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No | 33751331 | 37.21 | 33751331 | 37.21 |
| Yes | 56947429 | 62.79 | 90698760 | 100.00 |

1. What is the mean and standard deviation for depressive symptoms in 2012 and 2014?

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | N | Mean | Std Dev |
| |  | | --- | | **CESD\_12** | | **CESD\_14** | | |  | | --- | | 18207 | | 16043 | | |  | | --- | | 1.4245054 | | 1.3623722 | | |  | | --- | | 138.6543418 | | 137.0227489 | |

1. Is there a significant difference in depressive symptoms between men and women? Examine this for both the 2012 and 2014 waves.

**2012 Wave:**

Men mean depressive score = 1.2654

Women mean depressive score = 1.5561

Yes, t-value = -9.72 , p<0.0001. Men experience statistically significantly less depression in the 2012 wave.

**2014 Wave:**

Men mean depressive score = 1.1833

Women mean depressive score = 1.5075

Yes, t-value = -10.40, p<0.0001. Men experience statistically significantly less depression in the 2014 wave.

1. Is there a significant difference in depressive symptoms between those who are working vs. those who are not?

**2012 Wave:**

Employed mean depressive score = 1.0258

Unemployed mean depressive score = 1.7515

Yes, t-value = 24.56, p<0.0001. Employed experience statistically significantly less depression in the 2012 wave.

**2014 Wave:**

Employed mean depressive score = 0.9486

Unemployed mean depressive score = 1.6703

Yes, t-value = 23.30, p<0.0001. Employed experience statistically significantly less depression in the 2012 wave.

1. What is the N and correlation between depressive symptoms in 2012 and self-rated health in 2014?

**2012 Wave**

N = 16231, Pearsons Correlation = -0.38463

What is the N and correlation between depressive symptoms in 2014 and self-rated health in 2014?

**2014 Wave**

N = 16,031, Pearsons Correlation = -0.43169

1. How does the correlation change between 2012 and 2014, if at all? Why?

The weighted correlation from 2012 to 2014 became more negative, indicating a stronger negative relationship between depressive symptoms in 2014 when correlated with self-rated health in 2014 as opposed its correlation in 2012. The main reason I can think of the difference in correlations between these two time points is because overall mean depression scores dropped in 2014, indicating individuals are less depressed, thus the trend was likely magnified and the negative correlation became stronger.

1. What is the correlation between depressive symptoms in 2012 and the psychosocial scale you created?

**Impulse Scale (Higher scores = more impulsivity):**

Pearson’s Correlation = -0.17655

1. What is the correlation between depressive symptoms in 2014 and the psychosocial scale you created?

**Impulse Scale (Higher scores = more impulsivity):**

Pearson’s Correlation = -0.15071

1. How do scores on the psychosocial scale you created relate to self-rated health in 2014?

With weights incorporated, when running a simple linear regression model, using Impulse as a predictor for self-rated health at 2014, there is a positive relationship. Using self-rated health as the outcome variable, impulsivity scores are positively related to self-rated health:Y(self-rated health) = 2.430 + 0.180(Ximpulse). Our model was overall predictive of self-rated health F(1, 6156) = 130.64, P<0.0001. In total, this model explains 2.0% of the variance in self-rated health. Additionally, our simple linear model indicates that impulsivity is a statistically significant predictor of self-rated health, t(6156) = 35.29, p<0.0001.

Therefore, for every 1 unit increase in the impulse score, we expect to see a 0.180 in someone’s self-rated health score. This indicates that individuals with higher impulse scores rate there health at higher levels.

1. Does the relation between psychosocial scores and self-rated health change when you control for demographic variables (e.g., age, race, gender, education, coupleness)?

While adding weights and controlling for Age, race, whether coupled at 2014, gender, and degree, the relationship between Impulsivity & self-rated health remained positive and significant. Additionally, the overall multiple regression model, with the control variables remained significant in predicting self-rated health: F(6, 5712) = 111.31, p <0.0001. In total, this model explains 10.47% of the variance in self-rated health.

Additionally, when holding all control variables constant, Impulsivity remained a significant predictor of self- rated health with a beta of 0.15, t(5712) = 9.57, p < 0.0001. For every one unit increase in impulsivity, there is an expected 0.15 increase in someone’s health-rating.

Does using the sample weights change your results? If so, how? (Please describe.)

So overall, The weighting of the variables changes some characteristics of the descriptive statistics. For example. Our frequencies changed a lot (They are much larger). Additionally, along with this, our percentages changed as well – sometimes up to a ten percent difference (i.e our race/ethnicity variable changed by up to 10%!).

However, I could not help but notice this did not change the results of my T-test and correlation results. It slightly changed the regression results. Our parameter estimates changed slightly, however, I would argue not too much. With the sample weights included, we were able to explain slightly more variance in our models.