|  |  |
| --- | --- |
| proc means data=data2;  var Smokers Phys\_Inactive Excess\_Drinking Mental\_Stress Phys\_Stress Diabetic Insufficient\_Sleep Uninsured College Unemployed Housing\_Prob;  by Obese;  run; | |
|  |  |

The population has almost twice as many obese people than non-obese.

|  |  |
| --- | --- |
| proc glm data=data2 plot=diagnostics;  class Obese;  model Smokers Phys\_Inactive Excess\_Drinking Mental\_Stress Phys\_Stress Diabetic Insufficient\_Sleep Uninsured College Unemployed Housing\_Prob = Obese;  manova h = Obese / printe printh summary;  run; | |
| SMOKERS | |
|  |  |

The smokers distributions shows a little skew and with all observations having a Cook’s D under 0.0125, there are no influential points to be concerned about.

|  |  |
| --- | --- |
| PHYSICALLY INACTIVE | |
|  |  |

The physically inactive distribution is quite normal and all Cook’s D values are very low, under 0.008 so no influential points to worry about.

|  |  |
| --- | --- |
| EXCESS DRINKING | |
|  |  |

The excess drinking distribution is quite normal and all Cook’s D values are very low, under 0.005 so no influential points to worry about.

|  |  |
| --- | --- |
| MENTAL STRESS | |
|  |  |

The mental stress distribution is quite normal and all Cook’s D values are very low, under 0.004 so no influential points to worry about.

|  |  |
| --- | --- |
| PHYSICAL STRESS | |
|  |  |

The physical stress distribution exhibits a slight skew and all Cook’s D values are very low, under 0.01 so no influential points to worry about.

|  |  |
| --- | --- |
| DIABETIC | |
|  |  |

The diabetic distribution is fairly normal with a slight skew and all Cook’s D values are very low, under 0.008 so no influential points to worry about.

|  |  |
| --- | --- |
| INSUFFICIENT SLEEP | |
|  |  |

The insufficient sleep distribution is quite normal with a slight skew and all Cook’s D values are very low, under 0.007 so no influential points to worry about.

|  |  |
| --- | --- |
| UNINSURED | |
|  |  |

The uninsured distribution is fairly normal with a slight skew and all Cook’s D values are very low, under 0.008 so no influential points to worry about.

|  |  |
| --- | --- |
| COLLEGE | |
|  |  |

The college distribution is fairly normal with a slight skew and all Cook’s D values are very low, under 0.012 so no influential points to worry about.

|  |  |
| --- | --- |
| UNEMPLOYED | |
|  |  |

The unemployed distribution has some skew and possibly a couple influential points that will be reviewed.

|  |  |
| --- | --- |
| HOUSING PROBLEMS | |
|  |  |

The housing problems distribution has some skew and possibly a couple influential points that will be reviewed.

We remove 2 outliers for unemployed. We remove the data for Yuma county in Arizona due to the outlier it creates for unemployment. The county is along the Mexico border and is predominately a farming community with migrant (seasonal) workers. This situation is uncommon and not typical of U.S. counties. We also remove the data for Imperial county in California for the same reasons. It is adjacent to Yuma county.

We remove and 3 outliers for housing problems. We remove the data for Bethel, Northwest Arctic and Yukon-Koyukuk counties in Alaska for Severe Housing Problems. There are four factors that contribute to this category. They are housing units that lack complete kitchens, lack complete plumbing facilities, overcrowded, or severely cost burdened. These counties reside in Alaska where the cost to build is beyond what the residents can afford and therefore overcrowding is above normal compared to the rest of the United States. [Nathan Wiltse, Dustin Madden, 2018 Alaska Housing Assessment, Jan 17, 2018, https://www.ahfc.us/download\_file/view/5124/853]

|  |  |
| --- | --- |
|  |  |

After removing the outliers, there is no change in the fact that the population has almost twice as many obese people than non-obese.

|  |  |
| --- | --- |
| UNEMPLOYED | |
|  |  |

After removing the outliers, the unemployed distribution does not change. It still has some skew. All observations have a Cook’s D below 0.22 and this is deemed acceptable.

|  |  |
| --- | --- |
| HOUSING PROBLEMS | |
|  |  |

After removing the outliers, the housing problems distribution does not change. It still has a slight skew. All observations have a Cook’s D below 0.0125 and this is deemed acceptable.

We proceed with the dataset omitting the five data points.