**Question 7.2**

The REG Procedure

Dependent Variable: FVC\_father

|  |  |
| --- | --- |
| **Number of Observations Read** | 150 |
| **Number of Observations Used** | 150 |

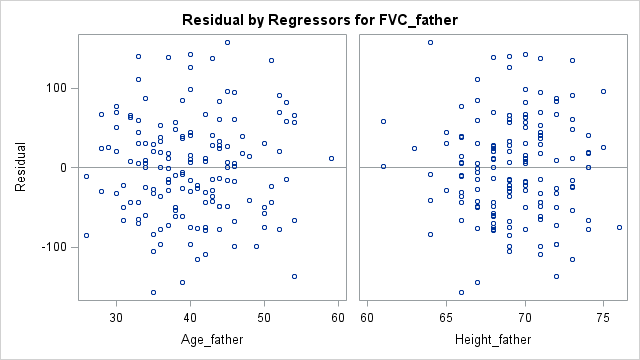
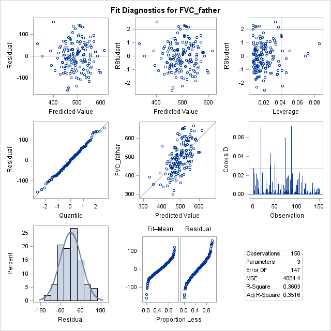
| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 338165 | 169083 | 41.40 | <.0001 |
| **Error** | 147 | 600404 | 4084.37907 |  |  |
| **Corrected Total** | 149 | 938569 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 63.90915 | **R2** | 0.3603 |
| **Dependent Mean** | 495.23333 | **Adjusted R2** | 0.3516 |
| **Coefficient Variance** | 12.90486 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -453.92042 | 135.96546 | -3.34 | 0.0011 |
| **Age\_father** | **1** | -2.77879 | 0.76109 | -3.65 | 0.0004 |
| **Height\_father** | **1** | 15.31441 | 1.88685 | 8.12 | <.0001 |

**Key Observations:**

* F Value: 41.4 and is Significant
* R2 = .3603
* Regression Line:
  + FVC\_father = -453.92042 + -2.77879 \* (Age\_father) + 15.31441 \* (Height\_father) + Error
* Residuals are mostly normally distributed with no apparent pattern
* There are a few observations that have high leverage. There are a few observations that have high influence.



**Question 7.2**

The REG Procedure

Dependent Variable: FVC\_father

|  |  |
| --- | --- |
| **Number of Observations Read** | 294 |
| **Number of Observations Used** | 294 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 1686.62825 | 562.20942 | 7.72 | <.0001 |
| **Error** | 290 | 21125 | 72.84634 |  |  |
| **Corrected Total** | 293 | 22812 |  |  |  |

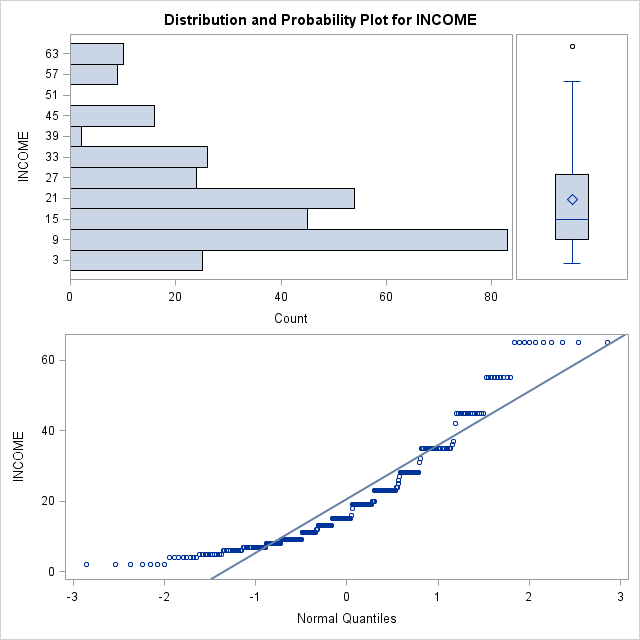
|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 8.53501 | **R-Square** | 0.0739 |
| **Dependent Mean** | 8.88435 | **Adj R-Sq** | 0.0644 |
| **Coeff Var** | 96.06784 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 12.44897 | 2.41928 | 5.15 | <.0001 |
| **INCOME** | **1** | -0.10321 | 0.03375 | -3.06 | 0.0024 |
| **SEX** | **1** | 1.82034 | 1.04396 | 1.74 | 0.0823 |
| **AGE** | **1** | -0.09894 | 0.02809 | -3.52 | 0.0005 |

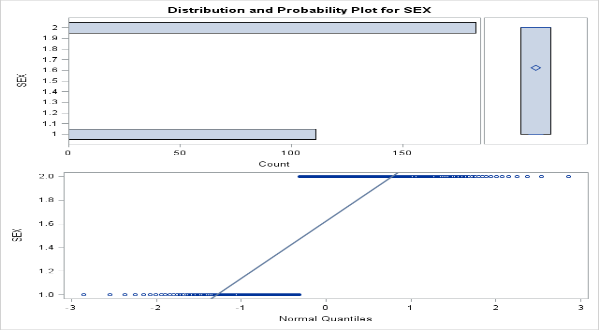
**Regression Line:**

CESD = 12.44897 + -0.10321 \* (INCOME) + 1.82034 \* (SEX) + -0.09894 \* (AGE) + Error

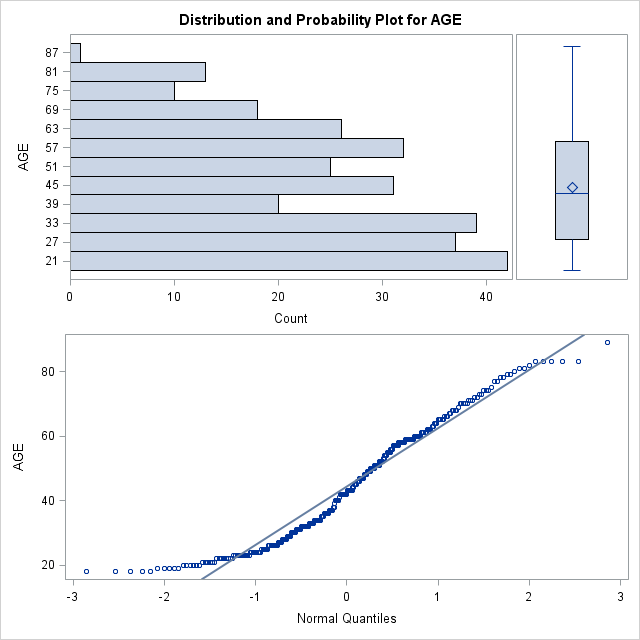
**Income**

From the QQ plot, we find that the data is abnormally bunched, where the low observations and high observations are not as extreme as they should be in a normal distribution. The residuals seem to be normally distributed.

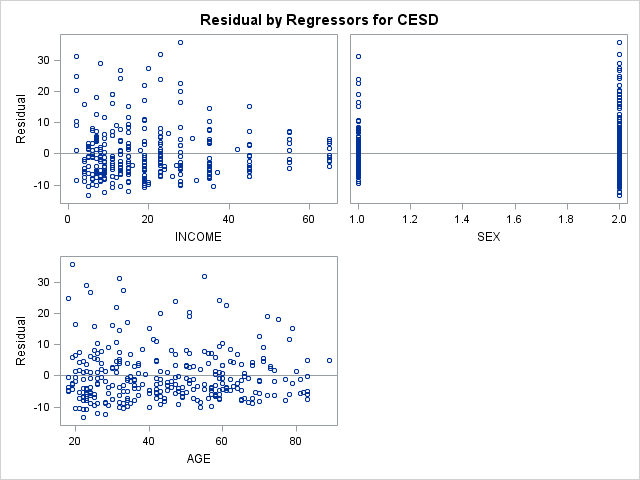
**SEX**

Given the binary nature of the data, it is hard to determine if it is normally distributed, but the QQ Plot seems to indicate that it is. Similarly, the residuals seem to be mostly evenly distributed

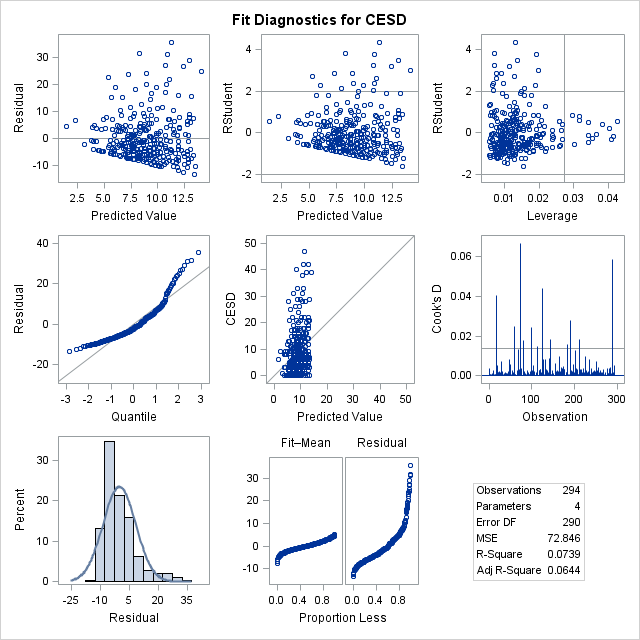
**AGE**



The QQ Plot shows the distribution is mostly normal, however, the lower value observations are not as extreme as a normally distributed sample would be. The plot of residuals are normally distributed



**Multivariable Correlation**



When plotting all three variables against the CESD, we find that the QQ Plot describes a data set that is clustered in the middle of the distribution, lacking the more extreme values on both ends that a normally distributed sample would have. Likewise, we find when plotting the residuals of each same against the mean of each Y at X, the residuals tend to be clustered lower than the mean, implying an uneven distribution.