**Question 6.2**

1. Perform a regression analysis of weight on height for fathers (and standardized regression equation)

The REG Procedure

Dependent Variable: Weight\_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** | 1 | 23222 | 23222 | 55.19 | <.0001 |
| **Error** | 148 | 62274 | 420.77118 |  |  |
| **Corrected Total** | 149 | 85496 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 20.51271 | **R-Square** | 0.2716 |
| **Dependent Mean** | 182.08667 | **Adj R-Sq** | 0.2667 |
| **Coeff Var** | 11.26535 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Weight\_father** | 150 | 182.08667 | 23.95408 | 27313 | 121.00000 | 245.00000 |
| **Height\_father** | 150 | 69.26000 | 2.77919 | 10389 | 61.00000 | 76.00000 |

|  |  |
| --- | --- |
| **Pearson Correlation Coefficients, N = 150  Prob > |r| under H0: Rho=0** | |
|  | **Height\_father** |
| **Weight\_father** | |  | | --- | | Correlation Coefficient: 0.52116 | | P Value: <.0001 | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | -129.02666 | 41.91228 | -3.08 | 0.0025 | 0 |
| **Height\_father** | **1** | 4.49196 | 0.60466 | 7.43 | <.0001 | 0.52116 |

**Normal Regression Equation:**

**Weight\_father** = -129.02666 + 4.49196 \* (**Height\_father)**

**Standardized Regression Equation:**

**Weight\_father** = 0.52116 \* (**Height\_father)**

**The null hypothesis of the slope = 0 (no correlation) is rejected. The Correlation Coefficient is 0.52116 with a P value of .0001, which supersedes almost all measures of level of significance.**

1. Perform a regression analysis of weight on height for mothers (and standardized regression equation)

The REG Procedure

Dependent Variable: Weight\_mother

|  |  |
| --- | --- |
| **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** | 1 | 14400 | 14400 | 16.60 | <.0001 |
| **Error** | 148 | 128380 | 867.43018 |  |  |
| **Corrected Total** | 149 | 142780 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 29.45217 | **R-Square** | 0.1009 |
| **Dependent Mean** | 146.97333 | **Adj R-Sq** | 0.0948 |
| **Coeff Var** | 20.03912 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Simple Statistics** | | | | | | |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Height\_mother** | 150 | 64.09333 | 2.46954 | 9614 | 57.00000 | 69.00000 |
| **Weight\_mother** | 150 | 146.97333 | 30.95568 | 22046 | 90.00000 | 267.00000 |

|  |  |
| --- | --- |
| **Pearson Correlation Coefficients, N = 150  Prob > |r| under H0: Rho=0** | |
|  | **Weight\_mother** |
| **Height\_mother** | |  | | --- | | Correlation Coefficient: 0.31758 | | P Value: <.0001 | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | -108.17272 | 62.66737 | -1.73 | 0.0864 | 0 |
| **Height\_mother** | **1** | 3.98085 | 0.97703 | 4.07 | <.0001 | 0.31758 |

**Normal Regression Equation:**

**Weight\_mother** = -108.17272 + 3.98085\* (**Height\_mother)**

**Standardized Regression Equation:**

**Weight\_mother** = 0.31758 \* (**Height\_mother)**

**The null hypothesis of the slope = 0 (no correlation) is rejected. The Correlation Coefficient is 0.31758 with a P value of .0001, which supersedes almost all measures of level of significance.**

1. Weight of mother = 267: Is this an outlier and if so, does it have a significant impact on the slope and therefore should be removed?

|  |  |  |  |
| --- | --- | --- | --- |
| **Obs** | **Weight\_mother** | **cookd** | **lev** |
| **7** | 206 | 0.026804 | 0.023462 |
| **42** | 241 | 0.036225 | 0.007571 |
| **45** | 231 | 0.095251 | 0.017197 |
| **94** | 260 | 0.069837 | 0.010667 |
| **104** | 186 | 0.027051 | 0.017197 |
| **107** | 234 | 0.093151 | 0.033161 |
| **115** | 220 | 0.035891 | 0.015964 |
| **130** | 130 | 0.027251 | 0.033161 |
| **144** | 267 | 0.056536 | 0.006676 |

**Above is the list of observations that should be removed from the analysis since they are significant influences on the regression slope. We see that the weight of 267 has a Cooks Distance of .057 and significantly affects the slope. It should therefore be removed.**

1. Why is the correlation higher for the father than the mother?

**If we take a look at the outliers for both, we find that the outliers for both fathers and mothers lie above the curve, and there are more outliers for mothers and the composite significance of those outliers on the regression line is greater.**

**Question 6.9**

**Each new point is an influential outlier as denoted by Cooks Distance**

**Without Additional Values**

**Income** = 27.773 + (-0.1621) \* **Age**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | 27.77274 | 2.32808 | 11.93 | <.0001 | 0 |
| **AGE** | **1** | -0.16206 | 0.04856 | -3.34 | 0.0010 | -0.19169 |

**Influential Outliers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **AGE** | **INCOME** | **cookd** | **lev** |
| **77** | 20 | 65 | 0.035546 | 0.009621 |
| **95** | 21 | 65 | 0.033938 | 0.009122 |
| **121** | 62 | 65 | 0.033216 | 0.006628 |
| **234** | 19 | 65 | 0.037206 | 0.010141 |

**Adding age= 42 income = 120**

**Income** = 28.21889 + (-0.16455) \* **Age**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | 28.21889 | 2.48924 | 11.34 | <.0001 | 0 |
| **AGE** | **1** | -0.16455 | 0.05194 | -3.17 | 0.0017 | -0.18199 |

**Influential Outliers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **AGE** | **INCOME** | **cookd** | **lev** |
| **77** | 20 | 65 | 0.030408 | 0.009605 |
| **95** | 21 | 65 | 0.029037 | 0.009106 |
| **121** | 62 | 65 | 0.028634 | 0.006619 |
| **234** | 19 | 65 | 0.031824 | 0.010125 |
| **295** | 42 | 120 | 0.065439 | 0.003450 |

**Adding age= 80 income = 150**

**Income** = 26.03193 + (-0.11268) \* **Age**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | 26.03193 | 2.61245 | 9.96 | <.0001 | 0 |
| **AGE** | **1** | -0.11268 | 0.05433 | -2.07 | 0.0389 | -0.12030 |

**Influential Outliers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **AGE** | **INCOME** | **cookd** | **lev** |
| **77** | 20 | 65 | 0.02899 | 0.009590 |
| **95** | 21 | 65 | 0.02761 | 0.009095 |
| **234** | 19 | 65 | 0.03041 | 0.010105 |
| **295** | 80 | 150 | 0.52119 | 0.016343 |

**Adding age= 180 income = 15**

**Income** = 26.95730 + (-0.14265) \* **Age**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | 26.95730 | 2.17992 | 12.37 | <.0001 | 0 |
| **AGE** | **1** | -0.14265 | 0.04449 | -3.21 | 0.0015 | -0.18411 |

**Influential Outliers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **AGE** | **INCOME** | **cookd** | **lev** |
| **77** | 20 | 65 | 0.033185 | 0.00881 |
| **95** | 21 | 65 | 0.031770 | 0.00838 |
| **121** | 62 | 65 | 0.029336 | 0.00596 |
| **234** | 19 | 65 | 0.034648 | 0.00925 |
| **295** | 180 | 15 | 0.097185 | 0.16334 |