Telco Regression Model

By Sondra Hoffman

July 18, 2021

Telco Regression Model

- The regression model indicates that each additional year in age is associated with an approximate $2.80 increase in income.

  Using the formula generated by the regression model, we can predict that a 27-year-old will make approximately $21.14.

Associations Analysis

1. This scatter plot shows a correlation between age and income.

2. There are a few outliers in the dataset.

3. A linear relationship exists between variables.

4. As a person gets older, there is a trend toward a higher income.

Associations of Age with Income
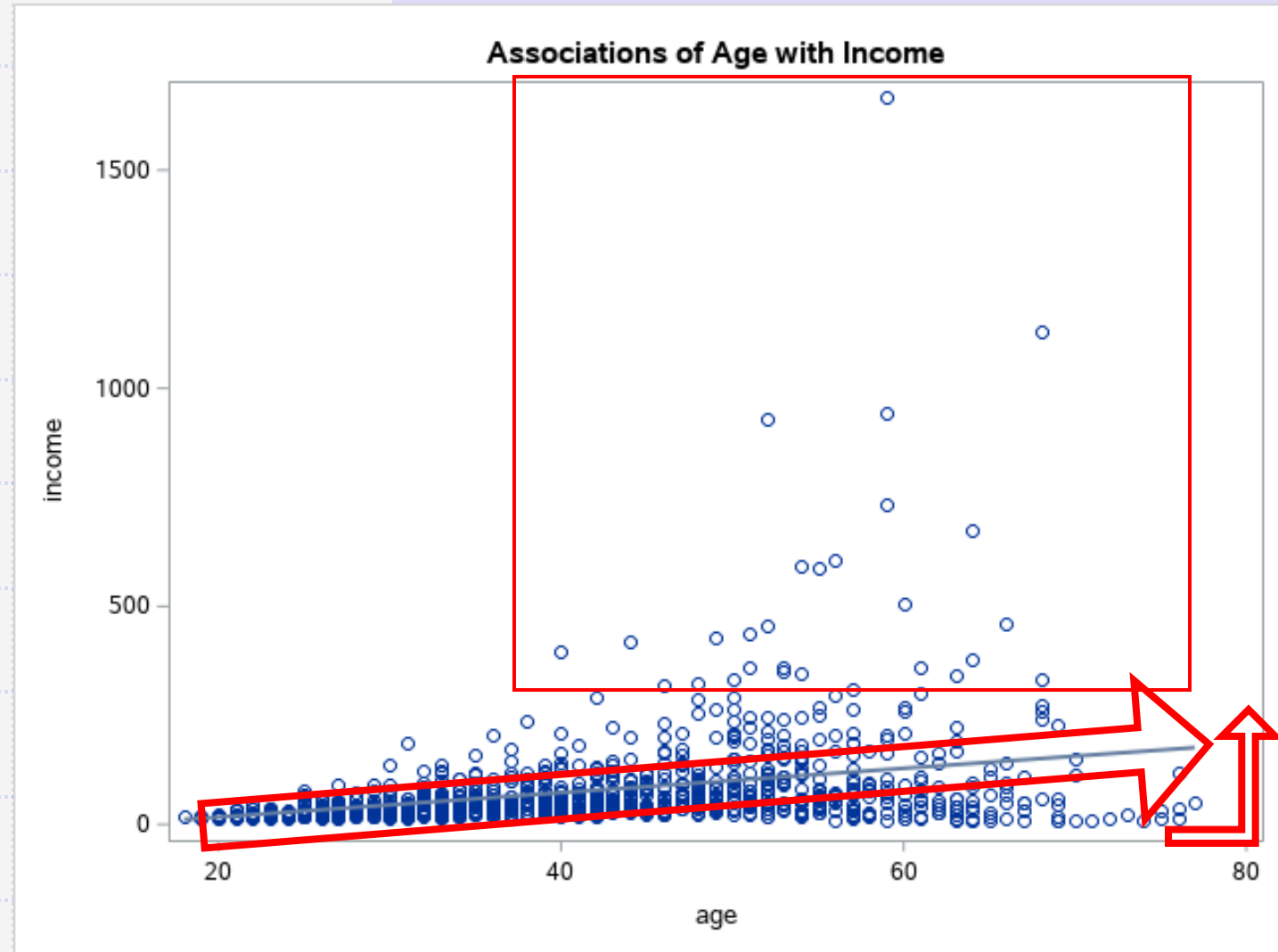
Correlations Analysis

1. Correlation tests for how well a continuous predictor explains a response in a regression analysis.

2. Due to the sample size, the correlation analysis shows a small p-value.

3. The correlation value of 0.3279 shows a positive correlation of the data because it is greater than 0.


Correlations and Scatter Plot with Age

2 Variables: age income

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
age	1000	41.68400	12.55882	41684	18.00000	77.00000	age
income	1000	77.53500	107.04416	77535	9.00000	1668	income

Pearson Correlation Coefficients, N = 1000
Prob > |r| under H0: Rho=0

age age	age 1.00000	income 0.32795 <.0001
income income	income 1.00000	age 0.32795 <.0001

Correlations and Scatter Plot with Age

Scatter Plot

Observations   1000
Correlation   0.3279
p-Value   <.0001
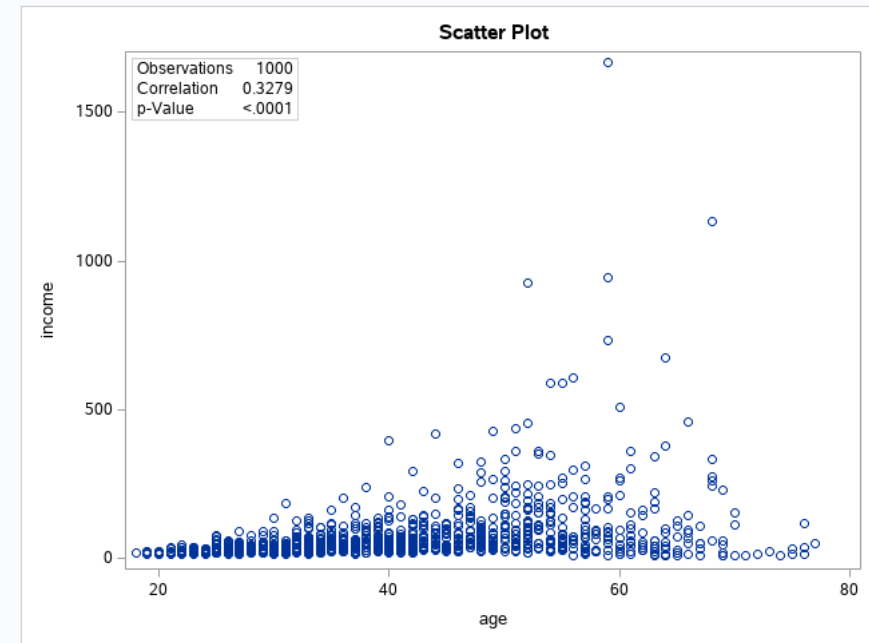
The REG Procedure
Model: MODEL1
Dependent Variable: income income

Number of Observations Read	1000
Number of Observations Used	1000

No missing data

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1231121	1231121	120.27	<.0001
Error	998	10215873	10236		
Corrected Total	999	11446995			

The simple linear regression model fits the data better than the baseline model.

Root MSE	101.17483	R-Square	0.1075
Dependent Mean	77.53500	Adj R-Sq	0.1067
Coeff Var	130.48924		

Age explains 11% of variation in the income variable.

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-38.98175	11.09583	-3.51	0.0005
age	age	1	2.79524	0.25488	10.97	<.0001

Regression equation is Income $\gamma = \beta_0 + \beta_1 x + \varepsilon$ or $y = (-38.98 + 2.80)$ age $+ \varepsilon$

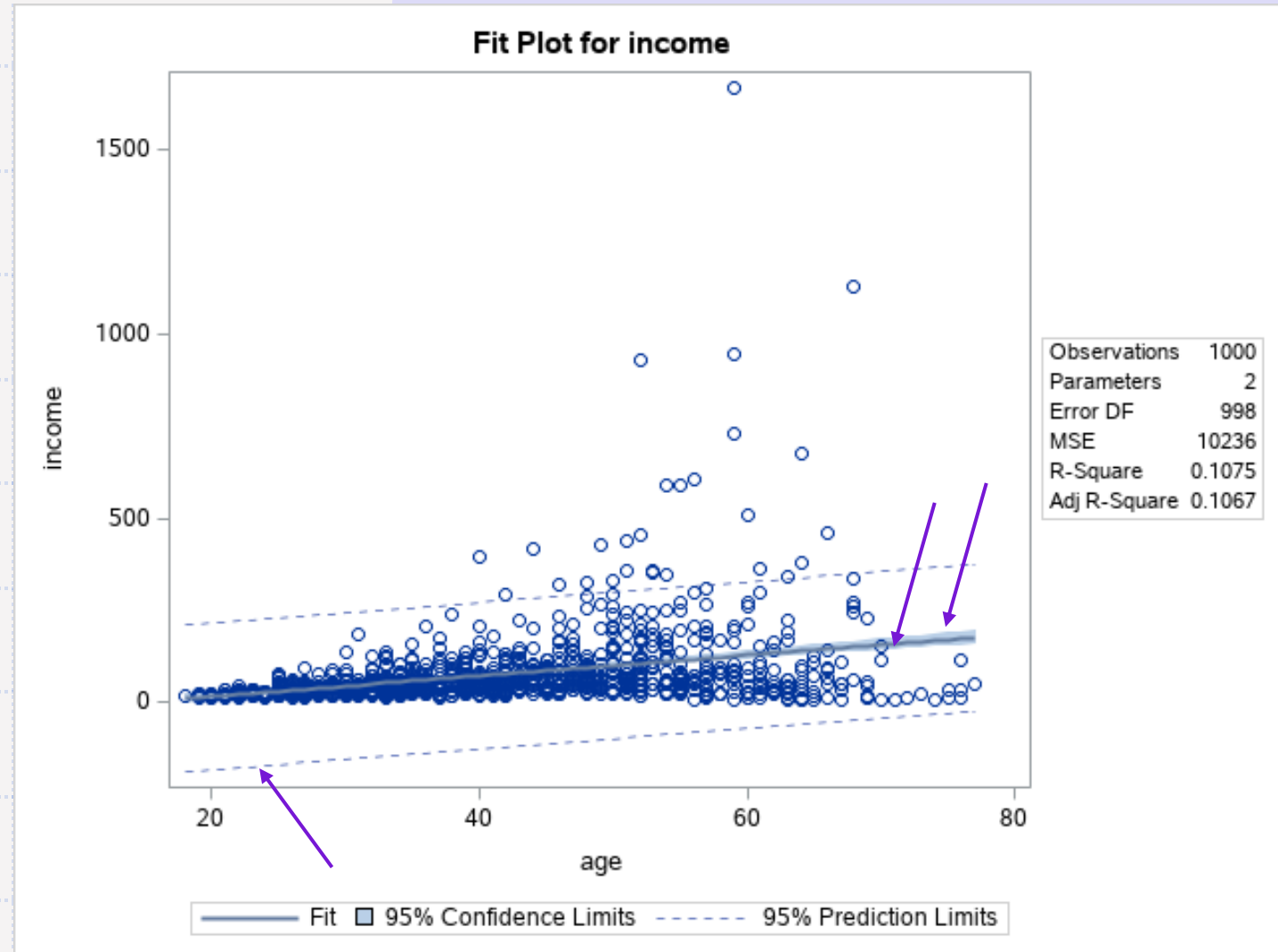This indicates a significant statistical difference in the slope from 0.

Fit Diagnostics for Income

1. This is a visual representation to support the assumptions in this model.

2. The Quantile graph shows minor deviations from normality.

3. The Residual graph shows there is a bell-shaped curve, supporting a normal distribution of data.

4. The Cook's D Observation shows a few outliers in the dataset.

Fit Diagnostics for income
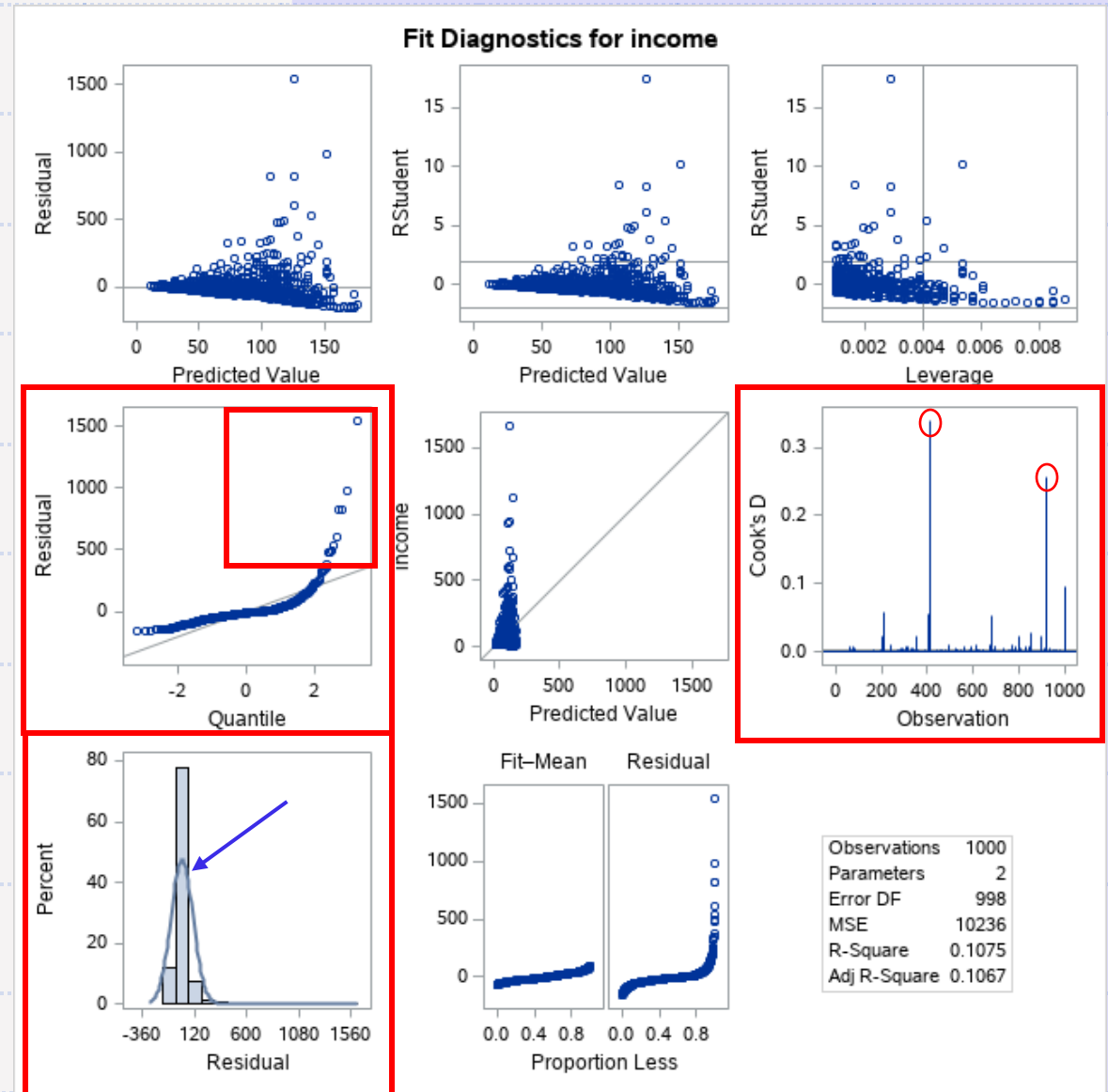
Fit Plot for Income

1. This shows a regression line placed over the data points in this dataset.

2. The shaded area around the regression line helps assess the precision of the mean around the income values.

3. The space between the dotted lines indicates a prediction limit.

Fit Plot for income

Observations	1000
Parameters	2
Error DF	998
MSE	10236
R-Square	0.1075
Adj R-Square	0.1067

Fit · 95% Confidence Limits · · · · 95% Prediction Limits

Conclusion

1. Telco data is sufficient to predict future incomes.
2. Each additional year in age is associated with a $2.80 increase in income.
3. A 27-year-old person at Telco can expect to make approximately $21.14.

References

SAS Institute, Inc. (2019). Statistics 1: Introduction to ANOVA, regression, and logistic

regression. https://vle.sas.com/course/view.php?id=2113

ZyBooks. (2018). Statistics in business analytics:

https://learn.zybooks.com/zybook/MIS445ZybookInteractiveText2019