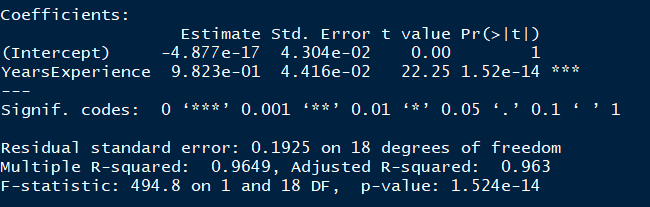
**Business problem**

To create a model that predicts us the estimated salary for an employee of a specific company based on the years of experience.

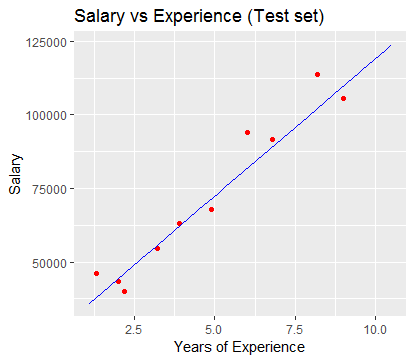
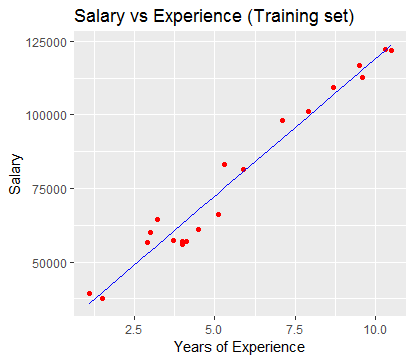
**Explaining the model**

1. Import the dataset, which is in a csv format.
2. After installing the package caTools, we’ve decided to split the sets on a 2/3 – 1/3 proportion for the training and test sets, respectively. Since we only have 30 entries on our dataset, 20 entries should be enough to train our model well enough.
3. After splitting, we could have transformed the categorical values to numerical, as the model cannot work with categorical values. However, the package we’ll use to fit the Regression Model will do just that, so it is not necessary to do it manually.
4. Next, the regressor is created. We want to see how the salary is impacted by the years of experience of the employees, so that means our dependent variable is Salary, and the independent is Years of Experience. The data we want to use in our regressor is our Training Set, because we’re training our model.
5. When we analyze our regressor, we can see that the coefficient of our independent variable is well below 0.05, which allows us to conclude (as suspected) that it has a strong effect on the dependent variable.



1. It’s now time to test our model; the first argument is the regressor we’ve created, and the second is the Test set.

**Analyzing the results**



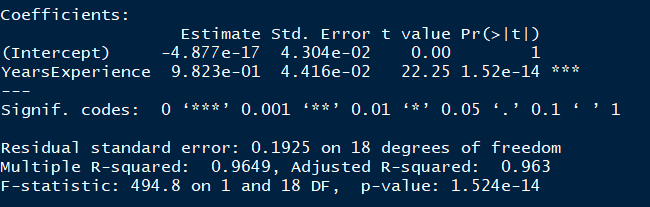
Being the red dots the real observations and the line our regressor, we can see that our model fits very well on the real observations. The results in both sets are similar, and according to our model, we have some individuals receiving a higher salary that they should and some other whose salary should be a little higher.

**Predictions**

If a person has 8 years of experience, our model predicts that it should have a salary of $100.513.

If a person has 20 years of experience, our model predicts that it should have a salary of $212.894.

**Evaluating the model’s performance**



Analyzing the Years of Experience variable, we can conclude that it has a positive effect on the dependent variable (the signal of the coefficient is positive), which means that as the Years of Experience increase, the Salary increases as well. Since it’s a Linear Regression model (with only one independent variable), we’ll look at the Multiple R-Squared, which is 0.9649 (the maximum value is 1, so we can conclude that it’s a robust model).