Interpreting Simple Regression Model (model1 in code)

Coefficients:

(Intercept) facebook

11.1740 0.2025

The estimated regression line is: Sales=11.174+0.2025\*Facebook+e

The intercept(b0) is 11.174 and can be interpreted as the predicted dollar sales value

for a Facebook advertising budget value of zero. So for a Facebook advertising budget equal

to zero, we can expect sales of 11.174 \*1000 = $11,174.

The regression coefficient(b1) shows that for a Facebook advertising budget equal to 1000 dollars,

we can expect an increase of 202.5 units (0.2025\*1000) in sales

i.e. Sales = 11.174 + 0.2025\*1000 = 56.44 units. This represents a sales of $213,670.

Summary of model1

Residuals:

Min -18.8766

1Q -2.5589

Median 0.9248

3Q 3.3330

Max 9.8173

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 11.17397 0.67548 16.542 <2e-16 \*\*\*

facebook 0.20250 0.02041 9.921 <2e-16 \*\*\*

Signif. codes: 0 '\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.13 on 198 degrees of freedom

Multiple R-squared: 0.332, Adjusted R-squared: 0.3287

F-statistic: 98.42 on 1 and 198 DF, p-value: < 2.2e-16

Residuals: As mentioned before, the residuals are the estimated coefficients denoted by ^. Residuals are the difference between the actual and estimated values.

The distribution of our residuals should ideally be symmetrical.

Coefficients: Our coefficients β0 and β1 represent the intercept and slope respectively. We have already interpreted these coefficients in the section above.

The coefficient standard error, as mentioned above, measures how much our coefficient estimates vary from the actual average value of our response variable.

In other words, it measures the accuracy of coefficient estimates. The closer our standard error is to zero, the better.

The coefficient t-value measures how far (in standard deviations) our coefficient estimate is from 0. A large t-value, relative to standard error, would provide

evidence against the null hypothesis and indicate that a relationships exists between the predictor and response variables. Predictors with low t-statistics can be dropped. Ideally, the t-value should be greater than 1.96 for a p-value to be less than 0.05.

The coefficient — Pr(>t) represents the p-value or the probability of observing a value larger than t. The smaller the p-value, the more likely we are to reject the null hypothesis. Typically, a p-value of 5% or less is a good cut-off point. Note the ‘Signif. Codes’ associated to each estimate, in our example. Three asterisks

represent a highly significant p-value.

Residual standard error: This measures the quality of our regression fit. It is the average amount the sales variable will vary from the true regression line.

Multiple R-squared: Besides the t-statistic and p-value, this is our most important metric for measuring regression model fit. R² measures the linear relationship between our predictor variable (sales) and our response / target variable (Facebook advertising). It always lies between 0 and 1. A number near 0 represents a regression that does not explain the variance in the response variable well and a number close to 1 does explain the observed variance in the response variable. In our example, the adjusted R² (which adjusts for degrees of freedom) is 0.3287 — only 32.87% of an increase in sales can be explained by Facebook advertising. If we perform a multiple regression, we will find that the R² will increase with an increase in the number of response variables.

F-statistic: This is a good indicator of whether there is a relationship between Y and X. The further our F-statistic is away from 1, the better our regression model. In our example, the F-statistic is 98.42, which is relatively larger than 1 given the size of our data set (200 observations). The F-statistic is more relevant in a multiple regression model.

So the four key indicators of model fit we would need to primarily focus on are the t-statistic, p-value, R² and F-statistic. The larger our t-statistic, the smaller the p-value. The smaller the p-value, the greater the odds of a relationship between X and Y. R² measures how well a model fits the data — if R² is close to 1, then this indicates that a large proportion of the variation in Y can be explained by X. The F-statistic shows the overall significance of the model. A large F-statistic will correspond to a statistically significant p-value (p < 0.05).