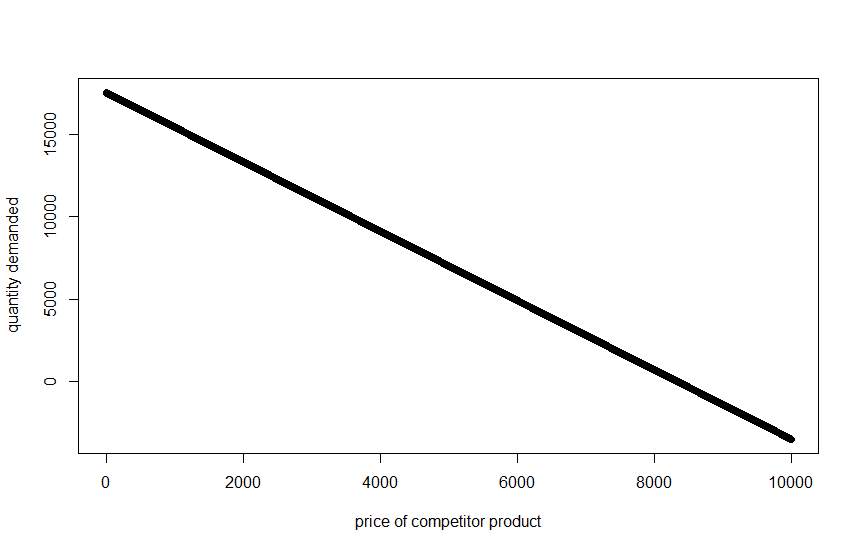
Question 1 – Interpreting Demand Curves :

1. The Klein Corporation’s marketing department, using regression analysis, estimates the firm’s demand function, the result being Q = -104 - 2.1P + 3.2I + 1.5A + 1.6Z Rsquared = 0.89 Standard error of estimate = 108 where Q is the quantity demanded of the firm’s product (in tons), P is the price of the firm’s product (in dollars per ton), I is per capita income (in dollars), A is the firm’s advertising expenditure (in thousands of dollars), and Z is the price (in dollars) of a competing product. The regression is based on 200 observations.
2. According to the statistical software, the probability is 0.005 that the t statistic for the regression coefficient of A would be as large (in absolute terms) as it is in this case if in fact A has no effect on Q. Interpret this result.

This is what is known as a p-value of the coefficient for advertising expense. This p-value essentially means that the statistic is significant (assuming 95% confidence) in the regression output (also assuming no collinearity of the variables, sometimes these values can be influenced by collinearity). Said another way, the probability of observing a value as or more extreme than those observed in the data given that the null hypothesis is true is 0.5%. At the end of the day, we can basically be assured that the coefficient (Ho: coeff = 0) does not equal 0.

1. If I = 5,000, A = 20, and Z = 1,000, what is the Klein Corporation’s demand curve? (Graph in R = keep in mind P means “Price”…) (SEE R CODE FOR DERIVATION)



1. If P = 500 (and the conditions in part b hold), estimate the quantity demanded of the Klein Corporation’s product.

**Q = 17,526 -2.1 \* 500 = 16,476**

1. How well does this regression equation fit the data given the Rsquared value? What additional metrics might be used to compare models?

The R-squared value of 0.89 is a pretty good fit to the data (assuming this is adjusted R-squared). This means that we are explaining 89% of the variation in quantity demanded with the variables included in the model. In order to compare models, we can look at metrics such as AIC, BIC, MSE, and forecast accuracy measured by MAPE. There are many more, but the most digestible approach for most executives is to hear how accurate the model is when generalizing to a previously unseen set of data. This allows executives to compare current approaches to new approaches while making a connection to the bottom-line impact of the accuracy of each model

Allen, W. Bruce; Weigelt, Keith; Doherty, Neil A.; Mansfield, Edwin. Managerial Economics: Theory, Applications, and Cases (Eighth Edition). (Page 125). W. W. Norton & Company. Kindle Edition.

Question 2 – Practical Learning: QSP – Compare statistical output to R:

1. Export CPS1985 data set into excel

library("AER")

data("CPS1985")

…

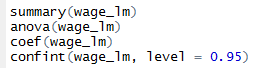
#Write data to excel

library(xlsx) #load the package

write.xlsx(x = CPS1985, file = "CPS1985\_DataDownload.xlsx",

sheetName = "BaseData", row.names = FALSE)

1. In R, create a linear model for wage by Union and Education. Hint: wage\_lm <- lm(wage ~ education, data = CPS1985). Then obtain these results on the model.



Summary results:

Call:

lm(formula = wage ~ education, data = CPS1985)

Residuals:

Min 1Q Median 3Q Max

-7.911 -3.260 -0.760 2.240 34.740

Coefficients:

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

(Intercept) -0.74598 1.04545 -0.714 0.476

education 0.75046 0.07873 9.532 <2e-16 \*\*\*

---

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

Residual standard error: 4.754 on 532 degrees of freedom

Multiple R-squared: 0.1459, Adjusted R-squared: 0.1443

F-statistic: 90.85 on 1 and 532 DF, p-value: < 2.2e-16

Anova:

Analysis of Variance Table

Response: wage

Df Sum Sq Mean Sq F value Pr(>F)

education 1 2053.3 2053.3 90.852 < 2.2e-16 \*\*\*

Residuals 532 12023.4 22.6

Coef:

(Intercept) education

-0.7459797 0.7504608

Conf interval:

2.5 % 97.5 %

(Intercept) -2.7997043 1.3077449

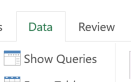
education 0.5957936 0.9051279

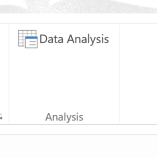
1. Perform same analysis using excel stat package. Are the results the same? If not, what specifically is different? Why does that matter?

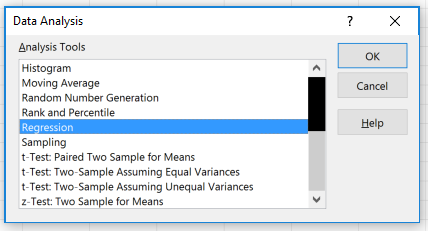
Overall. The results are very close. Although the output in R is rounded, we get to essentially the same answers. That being said, R only displays the coefficient out to 5 places after the decimal; however, it stores the entire string. In some cases, Excel will in fact output slightly different values. The problem with this output is that in some cases where the coefficients are small, rounded outputs can have a considerable impact on the resulting regression output. Therefore, it is not advisable to complete regressions in Excel.

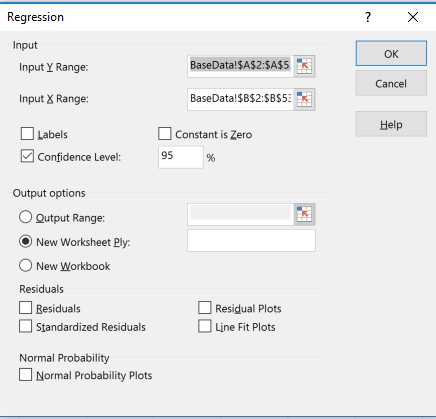
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.38192207 |  |  |  |  |  |  |  |
| R Square | 0.145864467 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.144258949 |  |  |  |  |  |  |  |
| Standard Error | 4.753986955 |  |  |  |  |  |  |  |
| Observations | 534 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 2053.290156 | 2053.290156 | 90.85197102 | 5.47384E-20 |  |  |  |
| Residual | 532 | 12023.40853 | 22.60039197 |  |  |  |  |  |
| Total | 533 | 14076.69868 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -0.745979668 | 1.045454059 | -0.713546101 | 0.475820796 | -2.799704259 | 1.307744923 | -2.799704259 | 1.307744923 |
| X Variable 1 | 0.750460751 | 0.078733726 | 9.53163003 | 5.47384E-20 | 0.595793611 | 0.905127892 | 0.595793611 | 0.905127892 |

Excel steps guide (PC):









Where data is structured in excel as….

