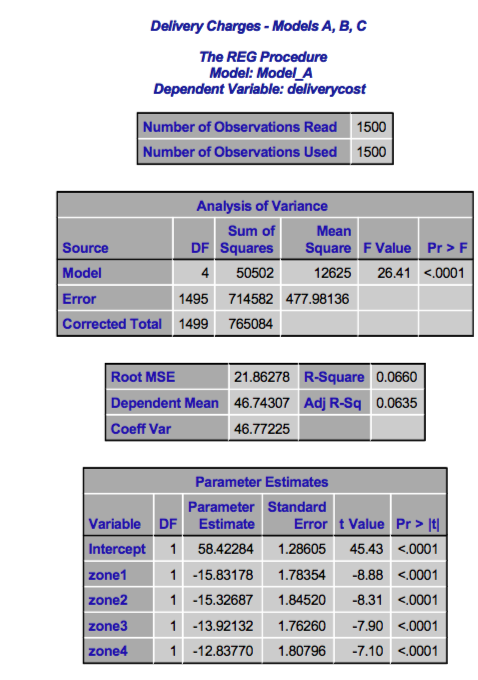
**Case: Estimating Invoiced Costs for Local Delivery Services**

**Answer 1 - a:**

**Model A:**



**Zone 1:** After accounting for the effects of zones 2, 3 and 4, this model shows that the average delivery cost for **zone 1 is $15.832 less** than what it is in zone 5.

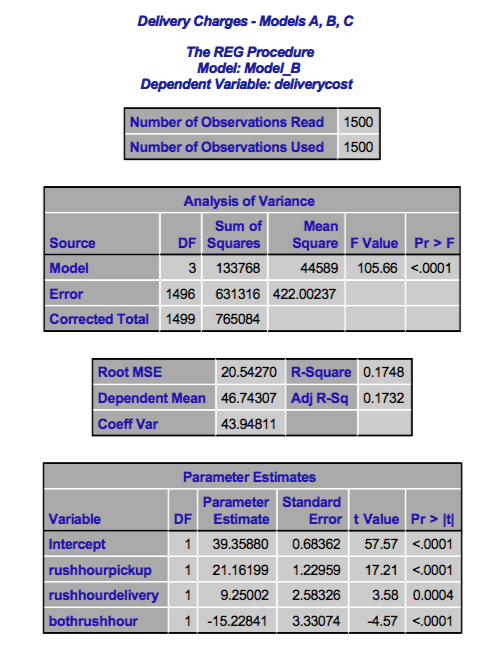
**Zone 2:** After accounting for the effects of zones 1, 3 and 4, this model shows that the average delivery cost for **zone 2 is $15.327 less** than what it is in zone 5.

**Zone 3:** After accounting for the effects of zones 1, 2 and 4, this model shows that the average delivery cost for **zone 3 is $13.921 less** than what it is in zone 5.

**Zone 4:** After accounting for the effects of zones 1, 2 and 3, this model shows that the average delivery cost for **zone 4 is $12.838 less** than what it is in zone 5.

**Answer 1 - b:**

**Model B:**



*Model B: Average Delivery Charges= f(rushhourpickup, rushhourdelivery, bothrushhour)*

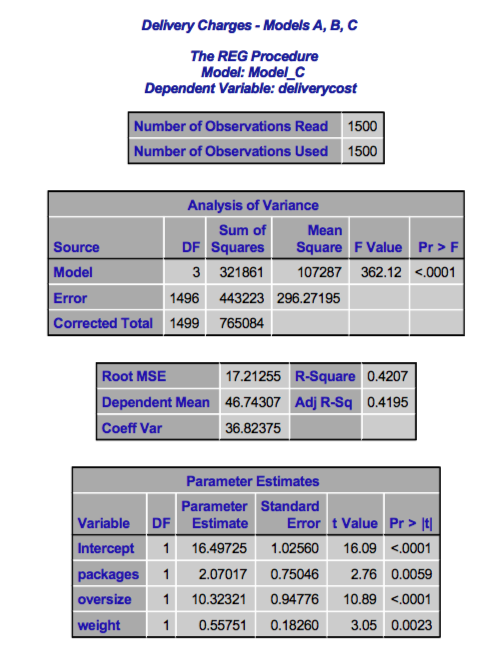
**Rushhourpickup:** If the package is picked up during Rush Hour then the average delivery cost increases by **$21.162** for each shipment.

**Rushhourdelivery:** If the package is delivered during Rush Hour then the average delivery cost increases by **$9.25** for each shipment.

**Bothrushhour:** If the package is picked up and is delivered during Rush Hours then the average delivery cost decreases by **$15.23** for each shipment.

**Answer 1 - c:**

**Model C:**



**Package:** This model shows that the Average delivery cost will increase by **$2.1** with an increase in every package.

**Oversize:** This model shows that the Average delivery cost will increase by **$10.32** if the shipment is oversized.

**Weight:** This model shows that the Average delivery cost will increase by **$0.56** with an increase of one kilogram for each shipment.

**Model D:**

**Zone 1:** After accounting for the effects of zones 2, 3 and 4, this model shows that the average delivery cost for **zone 1 is $15.65 less** than what it is in zone 5.

**Zone 2:** After accounting for the effects of zones 1, 3 and 4, this model shows that the average delivery cost for **zone 2 is $14.91 less** than what it is in zone 5.

**Zone 3:** After accounting for the effects of zones 1, 2 and 4, this model shows that the average delivery cost for **zone 3 is $13.85 less** than what it is in zone 5.

**Zone 4:** After accounting for the effects of zones 1, 2 and 3, this model shows that the average delivery cost for **zone 4 is $11.38 less** than what it is in zone 5.

**Rushhourpickup:** If the package is picked up during Rush Hour then the average delivery cost increases by **$17.25** for each shipment.

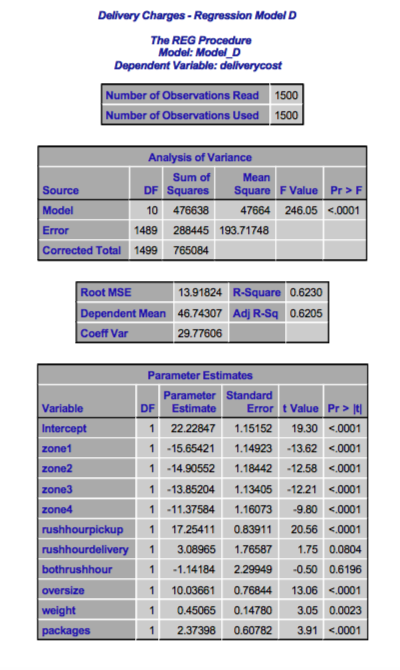
**Rushhourdelivery:** If the package is delivered during Rush Hour then the average delivery cost increases by **$3.09** for each shipment.

**Bothrushhour:** If the package is picked up and is delivered during Rush Hours then the average delivery cost decreases by **$1.14** for each shipment.

**Package:** This model shows that the Average delivery cost will increase by **$2.374** with an increase in every package.

**Oversize:** This model shows that the Average delivery cost will increase by **$10.04** if the shipment is oversized.

**Weight:** This model shows that the Average delivery cost will increase by **$0.45** with an increase of one kilogram for each shipment.



**Answer 2 – a:**

According to the Model A, Model C and the Model D the average delivery cost is similar when the transportation zone variables, the oversize indicator, the weight and the packages are considered. But when the values in Model B are compared with the values in model D, the average delivery cost does differ for the variables rush hour pick up, rush hour delivery and both rush hours. So, the estimates of the impact of individual factors depend on the rush hour variables in the model.

**Answer 3:**

According to Model D, the total weight of the shipment doesn’t contain significant predictive information on the margin after considering the effects of the other factors. This is because the R value of the weight variable in Model D is the same as the R value of the weight variable in Model C.

**Answer 4:**

According to Model D:

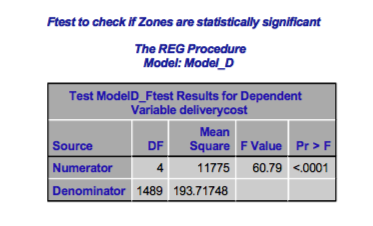
**Rushhourpickup:** If the package is picked up during Rush Hour then the average delivery cost increases by **$17.25** for each shipment but if the package is picked up and is delivered during Rush Hours then the average delivery cost decreases by **$1.14** for each shipment.

**Rushhourdelivery:** If the package is delivered during Rush Hour then the average delivery cost increases by **$3.09** for each shipment but If the package is picked up and is delivered during Rush Hours then the average delivery cost decreases by **$1.14** for each shipment.

**Answer 5:**

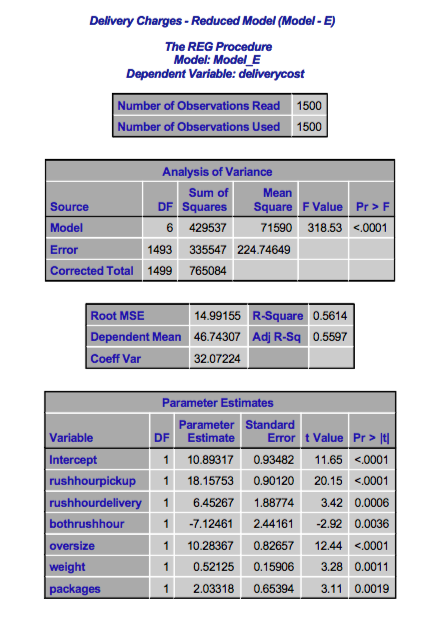
**F Test for Model D and Reduced Model E:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SSE | DF | MSE |
| Model E | 335546.5028 | 1493 | 224.7464854 |
| Model D | 288445.3306 | 1489 | 193.717482 |
| Difference | 47101.17212 | 4 | 31.02900348 |
|  |  |  |  |
| F Test | 60.78590797 | P Value | 1.31E-47 |

****

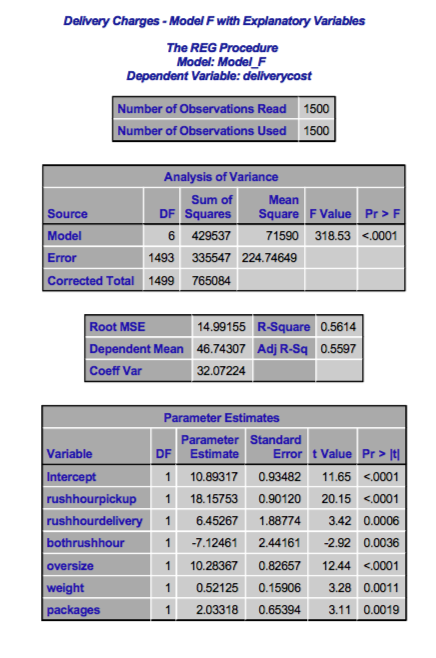
The F-Test between Model D and the reduced Model E shows that the **p-value is less than 0.0001**, which is less than the alpha value 0.05. This means even after considering the effects of the other variables, the variable indicators of transportation zones occur systematically making it statistically significant. The zone variables must be included in the regression model.

**Model E:**



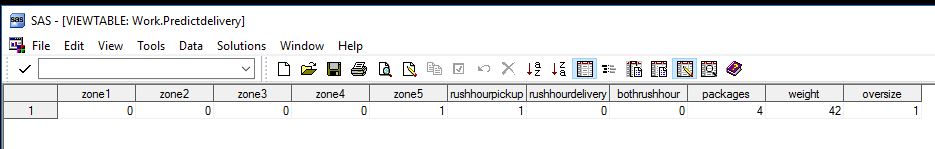
**Answer 6:**

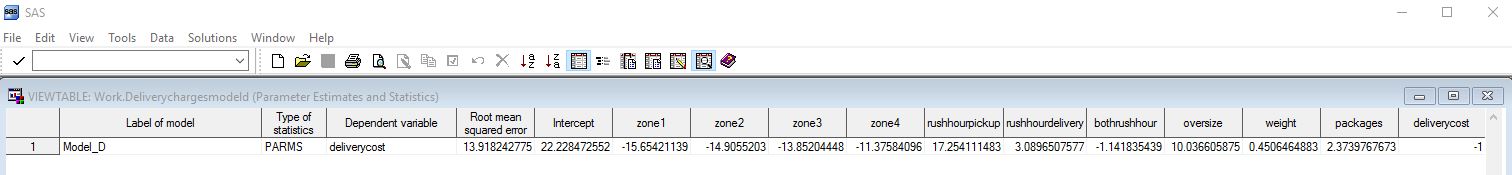
**Model F:**

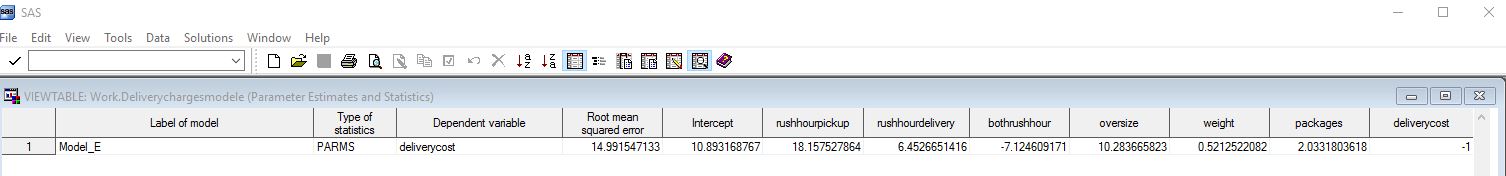


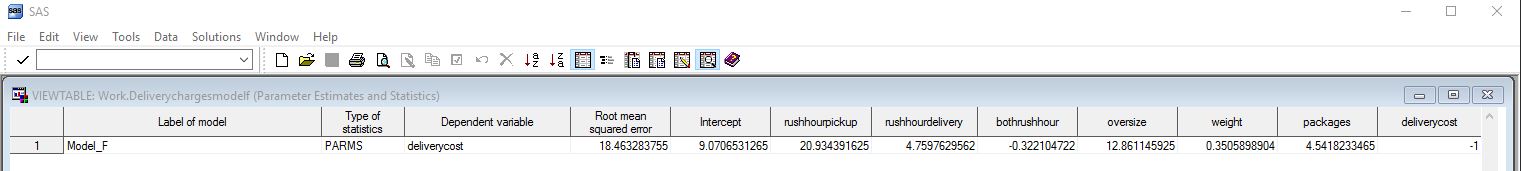
**Answer 7:**

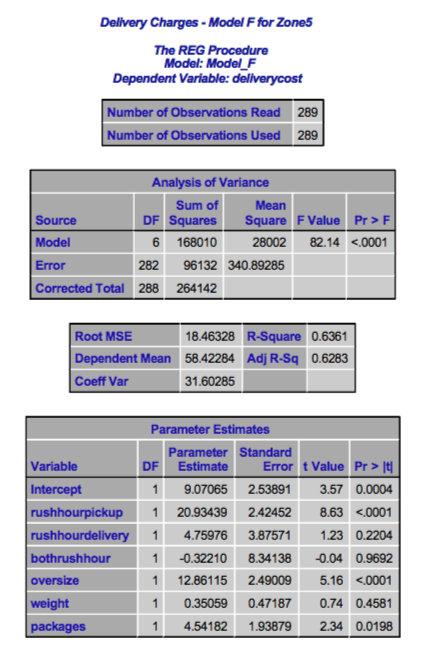
**The predicted cost of** **a delivery to Zone 5 of four packages weighing a total of 42 kg to be picked up at 8:30 AM and delivered at 10AM with oversized handling required:**









****

**Answer 8:**

**The average delivery cost in Model D = $22.23**

**The average delivery cost in Model E = $10.93**

**The average delivery cost in Model F = $9.07**

The average delivery cost in Model D is determined based on delivery zones, rush-hour indicators, oversized indicator, total weight of the shipment and number of packages, the average delivery cost in Model E is determined based on rush-hour indicators, oversized indicator, total weight of the shipment and number of packages, while we remove the zone indicators. The average delivery cost in Model F is determined using the same explanatory factors as Model E, but using deliveries to Zone 5 alone. Thus the average delivery cost differs in each model.

**Answer 9:**

**The Linear Regression Equation: Y = β0+β1X1 + β2X2 +β3X3 + ... + βkXk + ε**

Model D [Average Delivery Cost] = **f** (zone1, zone2, zone3, zone4, rushhourpickup, rushhourdelivery, bothrushhour, oversize, weight, number of packages)

**Model D – Y = 22.23 – 15.65\*(0) – 14.91\*(0) – 13.85\*(0) – 11.38 8(0) + 17.25(1) + 3.09\*(0) – 1.14\*(0) + 10.04\*(1) + 0.45\*(42) + 2.37\*(4)** = **$77.9**

Which explains that the average delivery cost for the hypothetical delivery in Zone 5 using Model D is **$77.9**

**Answer 10:**

When we consider the reduced models and the full models used based on the statistical significance, the reduced models would explain why they are considered an option. Consider **Model D,** which is the full model and the **Models A, B and C,** which are the reduced models.

The reduced models give us the significance of the variables. The **Model A** has a **p-value less than 0.0001,** which implies that the variables used in Model A are significant and occur systematically.

Similarly, the **Model B and Model C** have **p-value less than 0.0001,** which explains that the **variables used are highly significant.** Whereas, the **R Square** values of the **reduced models A, B and C are** **6.6%, 17.48% and 42.07%** respectively. This explains that these variables separately do not form a consistent model. While we consider the **full model D,** the **R Square** value is **62.3%** explaining that the full model is comparatively consistent and a significant model.

We consider the **reduced models** to understand the significance of a single variable or certain specific variables towards the model and the **full models** to understand the statistical significance of the model itself. Thus, allowing us to choose the best model and to know the significance of variable being used.

**Partitioned Data:**

**Advantage:** Explains the significance of the variables being used towards the model.

**Disadvantage:** Does not help to find a consistent and significant model.

**Augmented Data:**

**Advantage:** Explains the statistical significance of data for the whole model.

**Disadvantage:** Fails to determine the most significant variables to be considered or removed.