MGS 8040: Regression Analysis Exercise

**Consider the following output** of a regression model to predict ***home prices*** *($ thousand)* in a region based on *area of the house, age of the house, number of bedrooms* and *number of bathrooms*:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |
|  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |
|  |  |  |  |  |  |
| R Square | **0.6386** |  |  |  |  |
|  |  |  |  |  |  |
| Standard Error | **3.4876** |  |  |  |  |
| Observations | 20 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | **4** | 322.353 | 80.588 | **6.62561** | 0.002804 |
| Residual | **15** | **182.447** | 12.163 |  |  |
| Total | **19** | 504.800 |  |  |  |
|  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |  |
| Intercept | 215 | 3.5773 | 4.3141 | 0.0006 |  |
| Age | -1 | 0.3726 | -2.8058 | 0.0333 |  |
| Area | 0.05 | 0.0238 | 2.9589 | 0.0098 |  |
| Bedrooms | 9 | 2.0845 | 3.0827 | 0.0076 |  |
| Bathrooms | 13 | 1.0000 | 3.0000 | 0.0226 |  |

**Fill the shaded cells above.**

1. The **R-Square value** for the regression is equal to: **0.6386.** It means that 63.86% of the variation in home prices is accounted for by the variables in this model (Age, Area, Bedrooms, Bathrooms).
2. The **Standard Error** is equal to \_\_**3.4876**\_\_\_\_\_\_\_\_\_\_\_\_
3. Is the Regression significant at the 5% level**? Yes, Sig. F <0.05** At 1%? **Yes, Sig. F <0.01**
4. **The F value** for the regression is **6.62561**
5. The **coefficient -1.00 for age** means that **The price of a house goes down on average by $1000 for each year that it gets older, keeping Area, Bedrooms, and Bathrooms constant.**
6. Based on the regression above, the **predicted value** (in thousands of dollars) for a new home (age=0) that is 2000 sq. ft in area, with 3 bedrooms and 3 bathrooms is

**$381,000**

1. The **Margin of Error** for the 95% prediction interval for your prediction above is approximately

**$ 6,975** (two times the standard error, roughly).

1. You are testing the relationship between some X and Y for three different materials, A, B, and C. You wish to account for the material used in your regression analysis. Define the dummy variable values for a few observations of material type shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Y** | **X** | **Material** | **Dummy A** | **Dummy B** |
| 20 | 7 | A | **1** | **0** |
| 18 | 6 | B | **0** | **1** |
| 15 | 5 | B | **0** | **1** |
| 14 | 4 | C | **0** | **0** |
| 18 | 5 | C | **0** | **0** |
| 17 | 6 | A | **1** | **0** |
| 11 | 5 | A | **1** | **0** |