**Examples of exam questions**

Import the *dataset boston2.txt.* This dataframe contains information on crime rates in the districts of Boston. Next variables are available:

|  |  |
| --- | --- |
| **Name variabele** | **Description** |
| **obs** | Observation number |
| **crim** | Crime rate per district |
| **chas** | Variable which indicates whether the Charles River passes through the district   * Chas=1 : river passes the district * Chas=0: river does not pass the district |
| **rm** | Average number of rooms in a house |
| **roomfact** | Categorical variable indicating the type of house.   * low : average number of rooms < 6 * medium : average number of rooms =6 * high : average number of rooms > = 7 |
| **dis** | Average distance to the industrial area of Boston |
| **rad** | Average distance to the nearest highway |
| **zn** | Proportion of residential housing |
| **indus** | Proportion of industry in that district |
| **age** | Average age (in years) of the people living in that district |
| **crimcat** | * 1: if the crime rate is high * 0: if the crime rate is low |
| **international** | Variable indicating the type of district (this is a continuous variable). It describes the proportion of international residents in this district. |
| **internationfact** | Categorical variable indicating the type of district. This variable describes whether there are a lot of international residents in this districts.   * low : * medium : * high : * very\_high |

1. **Compare the average number of rooms in a house (rm) over the different types of districts.** Type of district is determined by the variable **internationfact** which gives an indication of the number of international residents in this district. Use significance level 0.05.

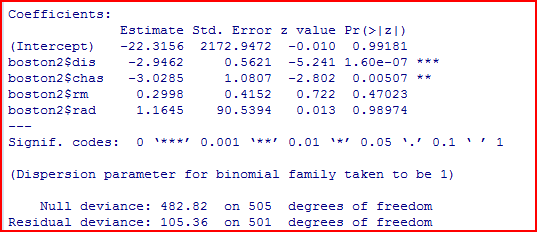
Answer this question in full detail in a structured way.

2.

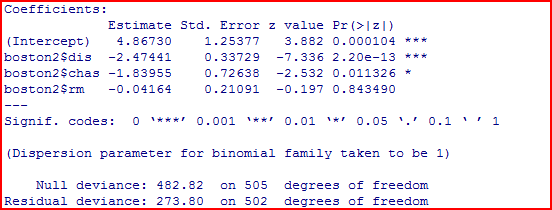
To answer this question, you have to use the **boston2** data again. Consider the simple linear regression model where we try to model the crime rate (crim) by the distance to the industrial area (dis). Predit the crime rate for a district where the distance to the industrial area is 3.123. Give also a 95 % confidence interval for the average crime rate you expect in that district.

1. We have used here again the **Boston2** dataset. For a description of the variables, we refer to question 1. We want to model why certain districts have a high crime rate and other districts have a low crime rate. We therefore use the **variable crimcat**. We give here the output of a few models.

Model 1:



Model 2:



Compare model 1 with model 2. Which one do you prefer. Explain why in a structured way?

1. We have used here again the **Boston2** dataset.

Check whether there is a significant difference in average crime rate whether the Charles river is passing or not (variable chas).

Be complete and structured in your answer.