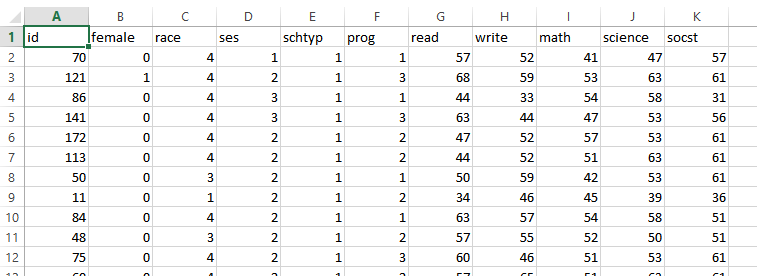
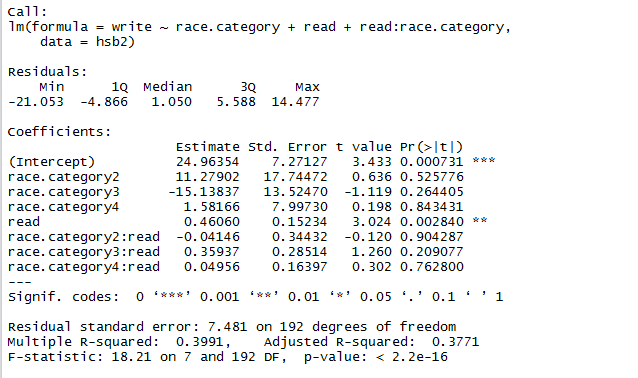
**MIDTERM 2 STAT 151 Version B**



This is part of a sample data containing 200 observations of high school students. It includes their gender (**female**), socio-economic status (**ses**), ethnic background (**race**), number of scores on standardized tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**). We are not using the other variables like schtyp or prog.

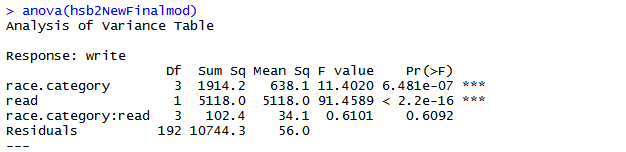
We fit the following model to the data using interactions:



Answer the questions (1-5) for the aforementioned linear model :

1. Write the fitted least square regression equation for the model presented above.
2. What is the slope coefficient for read given the race.Category2.
3. Write the hypothesis statement for the interactions between race and read
4. Given the above result are the interactions significant ? Give your rationale .
5. Is the read variable significant? Is any of the race.Category significant? Give your rationale.

Using the regressor variables race.category, read we fitted the model by the anova test to the data. Answer questions 6-10



1. Is the interactions between the regressors : race.category and read significant for the linear model with the response variable write. Give your rationale.
2. Are the regressors : race.category and read taken individually significant for the linear model with the response variable write. Give your rationale .
3. In the final model which regressor variables would you include? Give your rationale.
4. Prove: By matrix Algebra prove:



**Multiple Choice:**

1. If there is multicollinearity in the model :
2. The standard error of the slope coefficient decreases which causes the significant variable to become insignificant and therefore causes us to fail to reject the Null Hypothesis when we should reject the Null Hypothesis.
3. The standard error of the slope coefficient increases which causes the insignificant variable to become significant and therefore causes us to reject the Null Hypothesis when we should fail to reject the Null Hypothesis.
4. **The standard error of the slope coefficient increases which causes the significant variable to become insignificant and therefore causes us to fail to reject the Null Hypothesis when we should reject the Null Hypothesis**.
5. The standard error of the slope coefficient decreases which causes the insignificant variable to become significant and therefore causes us to reject Null Hypothesis when we should fail to reject the Null Hypothesis.
6. Gauss Markov theorem states that
7. If residual errors are independently distributed and they have zero expectation as well as constant variance. The least square estimator b of β is the most consistent and unbiased estimator. Therefore amongst all estimators of b the least square estimators has the smallest sampling variance (least mean squared error).
8. **If residual errors are independently distributed and they have zero expectation as well as constant variance. The least square estimator b of β is the most efficient and unbiased estimator. Therefore amongst all estimators of b the least square estimators has the smallest sampling variance (least mean squared error).**
9. If residual errors are independently distributed and they have zero expectation as well as constant variance. The least square estimator b of β is the most efficient and unbiased estimator. Therefore amongst all estimators of b the least square estimators has the largest sampling variance (largest least mean squared error).
10. If residual errors are independently distributed and they have zero expectation as well as constant variance. The least square estimator b of β is the most consistent and unbiased estimator. Therefore amongst all estimators of b the least square estimators has the largest sampling variance (largest least mean squared error).
11. If the Full model is significantly different and better than the null model then it makes sense that R1>>R0



1. Here q is the number of regressors included from the full model to obtain the Null Model.
2. Here q is the number of response variable omitted from the full model to obtain the Null or reduced Model.
3. **Here q is the number of regressors omitted from the full model to obtain the Null Model.**
4. None of the above.
5. Correlation and Independence:
6. **Independence implies lack of correlation but lack of correlation does not imply independence.**
7. Non Independence implies lack of correlation but lack of independence does not imply correlation.
8. Independence and correlation have no statistical connection.
9. None of the above
10. For linear Regression
11. Constant conditional variance of Y given X as the distribution of the response variable is not the same as the distribution of errors.
12. **Constant conditional variance of Y given X as the distribution of the response variable is the same as the distribution of errors.**
13. Constant conditional variance of Y given X as the distribution of the response variable might be the same as the distribution of errors.
14. None of the above
15. The normality constraint as a constraint for linear models :
16. If the normality constraint is not satisfied as a constraint for linear models we can conduct reliable Hypothesis especially for sample size less than 30.
17. We cannot determine the reliability of Hypothesis testing in linear models
18. **If the normality constraint is not satisfied as a constraint for linear models we cannot conduct reliable Hypothesis especially for sample size less than 30.**
19. None of the above