STAT 2131:
Applied Statistical Methods I
HW #2
Due Tuesday 11:00am, September 28th

1. You just started working for a company that is interested in looking at how radio advertising affected their sales. The data consist of the amount of sales Y_i in \$100 and the amount of radio advertising time X_i in hours for the ith month for $i = 1, \ldots, 24$.

   The following SAS code was run to obtain the output:

```
proc reg data= Sales;
model sales = radio /clb;
run;
```

```
---------------------------------------------------------------------------
                          The REG Procedure
                            Model: MODEL1
                       Dependent Variable: sales

                 Number of Observations Read          24
                 Number of Observations Used          24

                          Analysis of Variance
                          Sum of           Mean
Source              DF    Squares        Square         F Value    Pr > F
Model                1    1952.65488     1952.65488     712.51    <.0001
Error               22    60.29179       2.74054
Corrected Total     23    2012.94667

        Root MSE             1.65546       R-Square     0.9700
        Dependent Mean     114.88255       Adj R-Sq     0.9687
        Coeff Var            1.44100

                          Parameter Estimates
                 Parameter  Standard        t
Variable    DF    Estimate     Error     Value  Pr > |t|  95% Confidence Limits
Intercept    1  101.57570  0.60225    168.66    <.0001  100.32670    102.82470
Radio        1   1.15806   0.04338     26.69    <.0001   1.06809     1.24804
```

   (a) Write down the assumed model. Clearly state any assumptions and define all notations.

(b) Assume that the model in part (a) is appropriate. Provide a point estimate and a 95% confidence interval for the change in expected monthly amount of sales if monthly radio advertisement increases by 10 hours.

(c) Assume that the model in part (a) is appropriate. Provide a point estimate and a 95% confidence interval for the expected amount of sales if the radio advertisement time is zero hours.

(d) Perform a test to study whether radio advertisement tends to increase the amount of sales. State clearly the hypothesis, the testing statistic, the p-value and your conclusion.

2. A Psychiatrist studying the relationship between face memory function and schizophrenia disease collected the following variables for a random sample of 970 subjects: X_1 is the status of disease (1 for schizophrenia, 0 for healthy), X_2 is age in years, X_3 is gender (1 for female and 0 for male), and Y is the face memory score (a larger value means better performance). The data is stored in FACEMEM.xlsx.

   Assuming the linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ and normality of the error term, fit the model and answer the following questions. Use $\alpha = 0.05$ in all hypothesis testing.

   (a) What is the estimated face memory score for a 40-year-old healthy male?

   (b) One question of interest is whether schizophrenia people have different face memory compared to healthy people, controlling for age and gender. State the null and alternative hypotheses and conduct the hypothesis test.

   (c) What is the expected change in face memory if age increases by one year? Provide a point estimate and a 95% confidence interval.

   (d) Conduct an F-test for the overall significance of the model.

   (e) Is the model with three predictors significantly better than the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$? Conduct a statistical hypothesis test to answer this question.