Problem 1

We have a classical linear model

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where β_1 and β_2 are fixed parameters, u_i is a disturbance term that is independently and identically distributed with expected value 0 and population variance σ_u^2 and $i = 1, ..., n$ is the observation index. The OLS estimation helps us to obtain the following coefficient β_2 estimator:

$$\hat{\beta}_2 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{n\sum\limits_{i=1}^{n} X_i Y_i - \sum\limits_{i=1}^{n} X_i \sum\limits_{i=1}^{n} Y_i}{n\sum\limits_{i=1}^{n} X_i^2 - (\sum\limits_{i=1}^{n} X_i)^2}.$$

However, some algebraic transformations allow to show that

$$\hat{\beta}_2 = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}.$$

Prove this is true.
Hint: you can try to solve backwards (going from what you are given to prove to the OLS estimator)

Problem 2

A variable Y_i is generated as:

$$Y_i = \beta_1 + u_i$$

where β_1 is a fixed parameter, u_i is a disturbance term that is independently and identically distributed with expected value 0 and population variance σ_u^2 and $i = 1, ..., n$ is the observation index. The least squares estimator of β_1 is $\bar{Y}$, the sample mean of Y. However, a researcher believes that Y is a linear function of another variable X and uses ordinary least squares to fit the relationship:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X$$

calculating $\hat{\beta}_1$ as $\hat{Y} - \hat{\beta}_2\bar{X}$, where $\bar{X}$ is the sample mean of X. X may be assumed to be a nonstochastic variable. Determine whether the researcher's estimator $\hat{\beta}_1$ is biased or unbiased, and if biased, determine the direction of the bias.

Problem 3

The output below gives the result of regressing $FDHO$, annual household expenditure on food consumed at home, on EXP, total annual household expenditure, both measured in dollars, using the Consumer Expenditure Survey data set.

```
. reg FDHO EXP if FDHO>0

      Source |       SS       df       MS              Number of obs =     XXX
-------------+------------------------------           F(  1,  6332) = 3431.01
       Model |   972602566     1   972602566           Prob > F      =  0.0000
    Residual |  1.7950e+09  6332  283474.003           R-squared     =  0.3514
-------------+------------------------------           Adj R-squared =  0.3513
       Total |  2.7676e+09  6333   437006.15           Root MSE      =  532.42

------------------------------------------------------------------------------
        FDHO |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         EXP |   .0627099   .0010706    58.57   0.000     .0606112    .0648086
       _cons |   369.4418   10.65718    34.67   0.000     348.5501    390.3334
------------------------------------------------------------------------------
```

Unfortunately, some things are missing. Looking at this output, do the following tasks:

1. Give an interpretation of the coefficients estimations
2. Find the number of observations
3. Find the values of TSS, ESS and RSS
4. Explain in your own words what TSS, ESS and RSS are or provide formulas for them

Problem 4

We have a classical linear model

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where β_1 and β_2 are fixed parameters, u_i is a disturbance term that is independently and identically distributed with expected value 0 and population variance σ_u^2 and $i = 1, ..., n$ is the observation index.

Derive the OLS $\hat{\beta}_1$ and $\hat{\beta}_2$ estimators. Answers without a solution will not be accepted. You need to provide a full solution.