Question 1-(a)

[Answer] We need the assumption that the linear combination of $educ$, $fine$, and $prison$ have relevance to $marijuana$. This is reasonable at least in $fine$ and $prison$ as people fine or go to prison due to $marijuana$. According to $educ$, it has relevance to $log(earning)$ as higher education brings a lot of earning while it is orthogonal to other variables in linear equations. Besides their exogeneity, the linear regression has endogeneity caused by simultaneity from the feedback structure between $log(earning)$ and $marijuana$. In order consistently to estimate β_j, we need to suppose that $fine$, and $prison$ can remove the simultaneity from the linear regression as instruments.

Question 1-(b)

[Answer] First, we transform the second equation:

$$marijuana = \gamma_0 + \gamma_1 log(earnings) + \gamma_2 educ + \gamma_3 fine + \gamma_4 prison + u_2 \tag{1}$$

$$\rightarrow log(earning) = \frac{\gamma_0}{\gamma_1} - \frac{1}{\gamma_1} marijuana + \frac{\gamma_2}{\gamma_1} educ + \frac{\gamma_3}{\gamma_1} fine + \frac{\gamma_4}{\gamma_1} prison + \frac{1}{\gamma_1} u_2 \tag{2}$$

Then we can have the structural form of simultaneous linear regression equation. Then we run 1st stage linear regression like the following:

$$\widehat{marijuana} = \pi_1 + \pi_2 fine + \pi_3 prison \tag{3}$$

After we estimate $\widehat{marijuana}$, we apply $\widehat{marijuana}$ for removing the simultaneity from $marijuana$ in the original linear regression and run OLS on the following equation:

$$log(earning) = \beta_0 + \beta_1 \widehat{marijuana} + \beta_2 educ + u_1 \tag{4}$$

Now we can get the proper coefficients β_j after removing simultaneity.

Question 1-(c)

[Answer] Yes. For removing the simultaneity from $marijuana$, we use $fine$ and $prison$. As we use two factors, we do overidentification.

Question 2-(a)

[Answer] Suppose that the coefficient of $log(avgprc)$ is β_2. The sign of β_2 is minus. This means the negative correlation between $log(totqty_t)$ and $log(avgprc_t)$. This seems reasonable as the demand quantity is decreased as the fish price goes up or vice versa. The coefficient itself represents the price elasticity on demand as log is used on quantity and price. The elasticity consists of two things: substitute effect and income effect. For capturing income effect only, we have to see the substitutes of fish like meat or imported salmon. Even though the variance of β_2 is low compared to other coefficients, it is meaningless as the variance of the coefficients in dummy variables is large in general. On $\alpha = 0.05$, the t-test on H_0: $\beta_2 = 0$, H_1: $\beta_2 \neq 0$ can be done and H_0 is rejected as t-stat is significant ($\frac{-0.425-0}{0.176} = -2.41 < -1.96$). However, on $\alpha = 0.01$, we cannot reject H_0 as $-2.41 > -2.58$ and the t-test is insignificant. To see its significance among all variables including $log(avgprc_t$, we have to do f-test by comparing the critical value $\frac{(RRSS-URSS)/1}{URSS/(97-6)}$ with $F_{1-0.01}(1, 97-6)$.

Question 2-(b)

[Answer] From the quantity and price on supply and demand with simultaneity, we can have the following equations in structural form:

$$\begin{cases} q_i = \alpha_1 p_i + \beta_1 z_i + u_{1i} \\ q_i = \alpha_2 p_i + u_{2i} \end{cases} \tag{5}$$

where $E(z_i u_{1i}) = 0$. From its reduced form in equilibrium, we have:

$$\alpha_1 p_i + \beta_1 z_i + u_{1i} = \alpha_2 p_i + u_{2i} \tag{6}$$

Now we have:

$$p_i = \frac{\beta_1}{\alpha_2 - \alpha_1} z_i + \frac{1}{\alpha_2 - \alpha_1}(u_{1i} - u_{2i}) = \pi_p z_i + v_{pi} \tag{7}$$

$$Cov(p_i, u_{1i}) = Cov(\pi_p z_i + \frac{1}{\alpha_2 - \alpha_1}(u_{1i} - u_{2i}), u_{1i}) = \frac{\sigma_1^2 - \sigma_{12}}{\alpha_2 - \alpha_1} \neq 0 \tag{8}$$

$$Cov(p_i, u_{2i}) = Cov(\pi_p z_i + \frac{1}{\alpha_2 - \alpha_1}(u_{1i} - u_{2i}), u_{2i}) = \frac{\sigma_{12} - \sigma_2^2}{\alpha_2 - \alpha_1} \neq 0 \tag{9}$$

This means the linear regression with simultaneity has endogeneity. Therefore the estimates of the linear regression are wrong and the coefficients are wrong.

Question 2-(c)

[Answer] The two assumptions are the followings: First, $wave2_t$ and $wave3_t$ have no correlation on the error of the linear regression equation (instrumental validity, exogenous explanatory variables). Second, $wave2_t$ and $wave3_t$ have relevance on $log(avgprc_t)$. These assumptions are reasonable. The quantity of fish supply depends on the ocean wave height and they can shift the equilibrium points, which reveals the fish demand curve.

Question 2-(d)

[Answer] We do f test for checking the significance of two variables $wave2$ and $wave3$. F stat is:

$$\frac{(RRSS - URSS)/q}{URSS/(n-p)} = \frac{(15.576 - 10.934)/2}{10.934/(97-3)} = 19.915 \tag{10}$$

As $F_{1-0.05}(2, 94) = 3.09 < 19.915$, we reject the null hypothesis. The coefficients of the linear regression are not zero. This test result corresponds to the result of 1-(c) as $wave2_t$ and $wave3_t$ explains $log(avgprc_t)$ well.

Question 2-(e)

[Answer] From the first stage regression in 2-(d), we can have $log(\widehat{avgprc_t})$. Then we can run regression of the following:

$$log(\widehat{totqty_t}) = \beta_1 + \beta_2 log(\widehat{avgprc_t}) + \beta_3 mon_t + \beta_4 tues_t + \beta_5 wed_t + \beta_6 thrus_t \tag{11}$$

From the above regression, we can get $\beta = (8.164, -0.815, -0.307, -0.685, -0.521, 0.095)'$. $log(\widehat{argprc_t})$ removes the simultaneity part from $logavgprc_t$, which has the cause of endogeneity. For regressing, we have to use $log(avgprc_t)$, not $widehatlogprc_t$ for correct result.