**Question 7.1**

1. The interpretation of the model is as follows:
   * 74% of the variability in the starting salary of the sample is explained by GPA and METRICS
   * Both GPA and METRICS have a statistically significant effect on starting salary
   * A 0.1 increase in GPA leads to a $164.3 ceteris paribus increase in starting salary and folks who take Econometrics have a starting salary $5033 higher than those who don’t.
2. I would modify the equation as follows:

Where FEMALE=1, if female and 0 otherwise.

1. For this, I would need to add an interaction term FEMALE\*METRICS, changing the above equation to:

Together, (β4 + β5) show the difference between a Male taking Econometrics vs a Female taking Econometrics. If it is 0, we know the effect of taking Econometrics is the same. If not, we know there is some difference.

**Question 7.4**

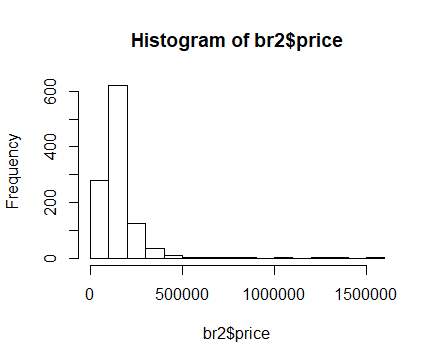
1. The coefficients of SQFT and AGE respectively are the change in house price values for a unit change in the value of SQFT and AGE respectively. For a 100 sqft increase in house area, the price increases by $7278.78 because it has a positive sign. Also, for a house that is older by 10 years, it’s price decreases by $1794.62 because its sign is negative. Both coefficients are statistically significant.
2. The estimated coefficients for variables D92 to D96 represent the change in the intercept value for years 1992 to 1996 respectively with respect to the year 1991. That means that all things remaining the same, a house price in the year 1993 differs from the price in 1991 by amount D93 and so on for other years.
3. I we had introduced a term for 1991, that would have introduced perfect collinearity because:

D91 + D92 + …. + D95 + D96 = 1

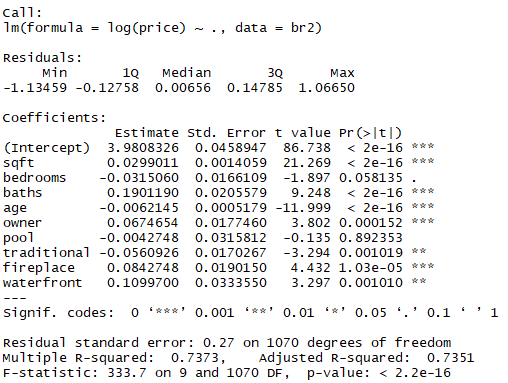
This would have led the OLS estimators to not work in this situation.

**Question 7.5**

1. The variables price, sqft, bedrooms, baths and age are heavily skewed and have a long right-tail. The histogram of price shows that as given below:



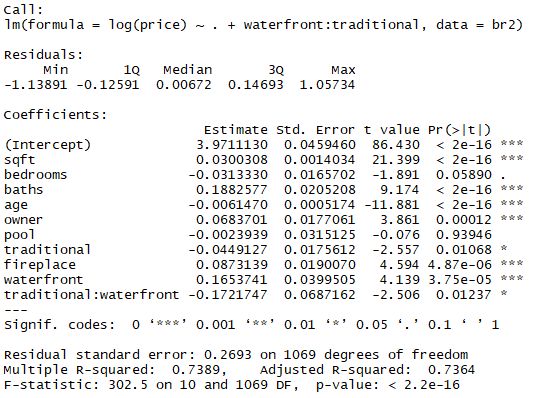
1. The constructed model summary output is below:



I would expect a positive sign for sqft, bedrooms, baths, pool, fireplace and waterfront. I would expect a negative sign for age, owner and traditional. However, some of the signs are a different from my estimates.

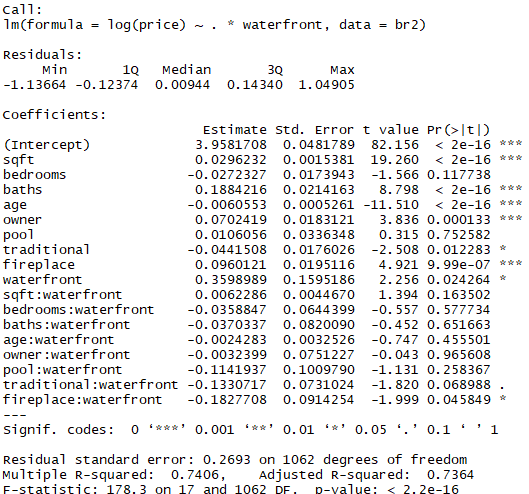
The interpretation of waterfront is the approximate percentage change in the price of a house if it is on the waterfront vs not being on the waterfront. The exact percentage change would be (eβwaterfront – 1).

1. The new model summary is below:



The effect of adding this variable on the adjusted R squared is minimal (0.7351 to 0.7364). This variable is significant at a 5% level. Now for its interpretation. This variable captures the interaction effect of being a traditional house by the waterfront, as compared to the effect of being a house that is traditional and is also at the waterfront.

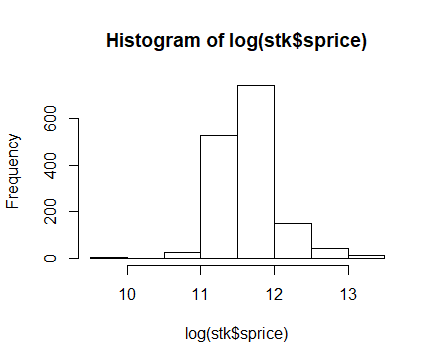
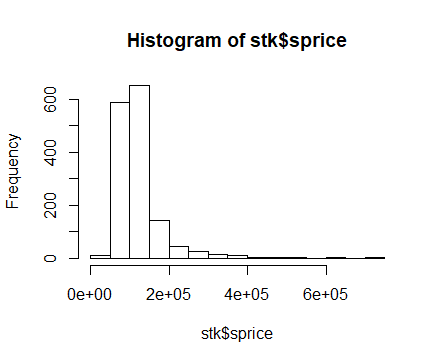
1. Comparing the previous 2 models with the model below, the p-value is greater than 0.05 in both cases, hence we cannot reject the Null. Thus, there is not enough evidence to say that non-traditional homes have a different regression model.



1. The prediction from the model in part c is $70,033,182 – astronomical.

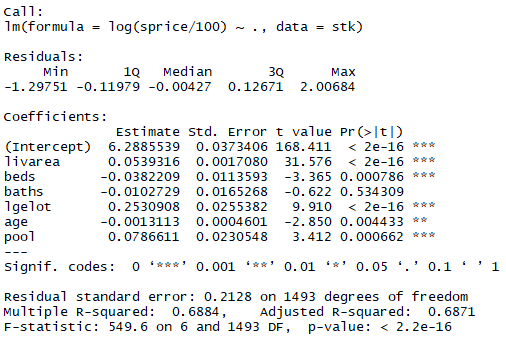
**Question 7.16**

1. Below are the histograms of sprice and ln(sprice)



Without logging, the distribution is heavily right-skewed. After logging, it resembles a normal distribution.

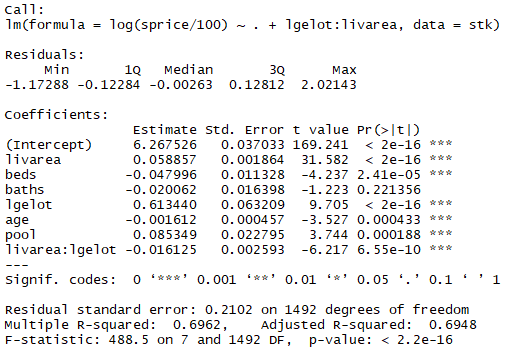
1. Below are the regression results:



This model has an adj R-sqrd of 0.6871, which is decent. The estimates are interpreted as below:

* LIVAREA: Statistically significant. Has a positive effect on sprice.
* BEDS: Statistically significant. Has a negative effect on sprice for some reason.
* BATHS: Statistically insignificant. Has a negative effect on sprice.
* AGE: Statistically significant. Has a negative effect on sprice.
* POOL: Statistically significant. Has a positive effect on sprice.

1. The indicator variable lgelot is statistically significant and appears to have nearly a 25% positive effect on the price of a house.
2. The model summary after introducing the interaction term is below:



The interaction term represents the extra change in price with respect to livarea, which comes in houses that have a large lot size. This appears to be negative, so it would imply that for the same increase in living area, a house with a small lot size would get a greater percentage increase in price than a price with a larger lot.

1. By performing a Chow-test of a model without the intercept and slope indicator variables, vs one with the indicator variables, I can reject the Null that the models for houses with large lots and those without are equivalent. The indicator variable is significant at least one of the place in the intercept or interaction terms.