1. **The house was sold for $37,704 more than what was predicted by the model. The selling price for this home is 37,704 more than we would have predicted based on the multiple regression model. It would appear that there is some other factors outside the size of homes and size of the lot that affect the sale price of the home.**
2. **Home selling price is predicted to increase $53.8 for every 1 square foot increase holding lot size constant. For each square-foot unit increase in home size, the house selling price would increase by 53.8 dollars. This is the interpretation of the slope.**
3. **Given that the y intercept and the lot size is fixed, the only variable left is home size.**
4. **Interpretation: In order for the lot size to have the same impact as a one-squared-foot increase in home size, the lot size would have to increase by 18.84 square feet.**
5. **Thousands of dollars would be the new unit of measurement for y and yhat as opposed to dollars so the partial coefficients would need to move three decimal places to adjust for the new unit of measurement. It would not change how the selling price is calculated, it just changes what kind of units it is measured by. Changing from 1 to 1,000 would require us to move the decimal three space to the left for the coefficients for size of home, and size of lot. In other words, the original equation has the A, B1 and B2 presented in dollars so the yhat output from the model is also in dollars. To state the yhat in thousands of dollars the A, B1 and B2 have to be reduced by 1,000.**

# If house selling prices are changed from dollars to thousands of dollars, it means that each part in the equations is divided by 1000. The new intercept = -10,536/1000 = -10.536; β1 = 53.8/1000= 0.0538; β2 = 2.84/1000 = 0.00284. Therefore, if house selling prices are changed from dollars to thousands of dollars, the prediction equation: y hat = -10.536 + 0.0538X1 + 0.00284X2.

1. **It is a global F test that tells us how likely it is that any of the explanatory variables have an effect on the percentage of adult residents who are registered to vote. We are trying to disprove the null hypothesis that the explanatory variables tested do not have an effect on y.**
2. **Unlike the F test, the t test determines if a specific variable has an effect on the percentage of adult residents who are registered to vote while holding all other variables constant.**
3. **The p value of the t-test for x1 is 0.03 which is significant at the 0.05 level. This leads us to conclude that the percentage of adult residents owning homes likely does have an effect on the percentage of adult residence who are registered to vote.**

**“Test of H0: B1 =0; P = 0.03, so there is considerable evidence that X1 has an effect on y, controlling for the other Xs.”(from the book)**

**The T statistic next to X1 represents the statistical significance of the variable in explaining the variation of y while holding the other variables constant.**

1. **meaning the total person trips would equal 51.6 if both linear distance to CBD and number of autos owned were 0.**

**For every unit increase in distance to CBD, there is a –10.12 unit change in total person trips, holding all else constant.**

**For every unit increase in number of autos owned, there is a 7.15 unit change in total person trips, holding all else constant.**

**The standard error of slope for linear distance to the CBD is 5.976. The large amount of variability, the less accurate of the model.**

**The standard error of slope for number of auto owned is 0.667. The small amount of variability, the more accurate of the model.**

**R2 is 0.965. It means that 96.5% of total person trips are explained by distance to the CBD and number of auto owned.
Adjusted R2 is 0.9563 < 0.965. The difference between Adjusted R2 and R2 is 0.0087. The difference is small. It means that adding more independent variable in the model help to add explanatory power.**

1. **In the multivariate model, the p-value of the f statistics = 1.5e-6 which is less than α = .01. Therefore, we reject the null, there is a relationship between the independent variables and the dependent variable.**
2. The interpretation:
* The final regression used in model 6 because the adjusted R-squared is 0.1747 which is between 0.15 and 0.2 and all the independent variables are significant.
* The intercept is 1.589650. It means that if the independent variables are equal 0, so health index for non-smokers is 1.589650.
* If all other independent variables are constant, health index for smokers is 0.265159 compared to non-smokers.
* If all other independent variables are constant, health index for whose parents are educated is -0.018235. This means that it is expected that their health will be worsened/decreased by 0.018235 if their parents are not educated.
* If all other independent variables are constant, health index depending on social class varies by (-0.074075) on a social hierarchical class.
* If all other independent variables are constant, for additional one unit in happiness scale, you can expect health index for smokers to increase by (0.279124).
* If all other independent variables are constant, health index depending on the number of hours spend on TV increases by (0.066850)
* If all other independent variables are constant, health index depending on the LeanDem decreases by (-0.148676). You can expect a decrease in the health index by (-0.148767) for additional one unit in (leanDem).
* P-values of each variable is less than 0.1, so we can conclude that each independent variable is significant.

SE = 0.231299 , it means that the standard error of slope is 0.231299 The small amount of variability, the more accurate the model.

### P-values of model is 2.001e-12 <0.1, so we can conclude that there is at least one variable which has relationship to health index.

Adjusted R-squared is 0.1747 . It means that 17.47% of health index is explained by smoke, parents’ education, class, happiness level, and TV hours.