Jeffrey Chang
jlc190004

EXERCISE 7.1:

An economics department at a large state university keeps track of its majors' starting salaries. Does taking econometrics affect starting salary? Let SAL = salary in dollars, GPA = grade point average on a 4.0 scale, METRICS = 1 if student took econometrics, and METRICS = 0 otherwise. Using the data file *metrics.dat*, which contains information on 50 recent graduates, we obtain the estimated regression

SAL = 24200 + 1643* *GPA* + 5033* *METRICS* R^2=0.74
(se)   (1078)   (352)   (456)

(a) Interpret the estimated equation.

- GPA: When increased by 1, and other variables remain the same, we can estimate that the average starting salary is estimated to increase $1643

- METRICS: Students who did not take econometrics are estimated to have a starting salary lower by $5033 than students who take econometrics

- 24200: This is the intercept, which suggests GPA = 0 and not taking econometrics class, for someone's starting salary is $24,200. But everyone somehow will have a non-zero GPA, so this looks unrealistic.

- R^2: 74% of the variation of starting salary is explained by GPA and METRICS

(b) How would you modify the equation to see whether women had lower starting salaries than men? (Hint: Define an indicator variable FEMALE = 1, if female; zero otherwise.)

According to the *metrics.dat*, I use gender to modify the equation.

SAL = β_1 + $\beta_2 GPA$ + $\beta_3 METRICS$ + $\beta_4 FEMALE$ + e

So, define an indicator variable if FEMALE = 1 and if FEMALE = 0:

SAL = β_1 + $\beta_2 GPA$ + $\beta_3 METRICS$, if FEMALE = 0

SAL = $(\beta_1 + \beta_4)$ + $\beta_2 GPA$ + $\beta_3 METRICS$, if FEMALE = 1

(c) How would you modify the equation to see if the value of econometrics was the same for men and women?

I add the new variable to present the relation between gender and econometrics, and the modified equation is:

SAL = β_1 + $\beta_2 GPA$ + $\beta_3 METRICS$ + $\beta_4 FEMALE$ + β_5 *METRICS* × *FEMALE* + e

So, define an indicator variable if FEMALE = 1 and if FEMALE = 0:

SAL = β_1 + $\beta_2 GPA$ + $\beta_3 METRICS$, if FEMALE = 0

SAL = $(\beta_1 + \beta_4)$ + $\beta_2 GPA$ + $(\beta_3 + \beta_5)$ *METRICS*, if FEMALE = 1

Jeffrey Chang
jlc190004

EXERCISE 7.4:

In the file *stockton.dat* we have data from January 1991 to December 1996 on house prices, square footage, and other characteristics of 4682 houses that were sold in Stockton, California. One of the key problems regarding housing prices in a region concerns construction of ''house price indexes,'' as discussed in Section 7.2.4b. To illustrate, we estimate a regression model for house price, including as explanatory variables the size of the house (SQFT), the age of the house (AGE), and annual indicator variables, omitting the indicator variable for the year 1991.

PRICE = β_1 + $\beta_2 SQFT$ + $\beta_3 AGE$ + δ_1D92+ δ_2D93 + δ_3D94 + δ_4D95 + δ_5D96 + e

The results are as follows:

Stockton House Price Index Model

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	21456.2000	1839.0400	11.6671	0.0000
SQFT	72.7878	1.0001	72.7773	0.0000
AGE	−179.4623	17.0112	−10.5496	0.0000
D92	−4392.8460	1270.9300	−3.4564	0.0006
D93	−10435.4700	1231.8000	−8.4717	0.0000
D94	−13173.5100	1211.4770	−10.8739	0.0000
D95	−19040.8300	1232.8080	−15.4451	0.0000
D96	−23663.5100	1194.9280	−19.8033	0.0000

(a) Discuss the estimated coefficients on SQFT and AGE, including their interpretation, signs, and statistical significance.

SQFT: *(No units found, so assume SQFT 's unit is K)*
- Add 1K square footage will increase the house price by $72.79, when other factors fixed.
- Expectation of bigger house will have higher price, so is the positive estimated coefficient
- It is statistical significance different from zero.

AGE:
- Add 1 more year will decrease the house price by $179.46, when other factors fixed.
- Expectation of older house will have lower price, so is the negative estimated coefficient
- It is statistical significance different from zero.

Jeffrey Chang
jlc190004

(b) Discuss the estimated coefficients on the indicator variables.

The estimated coefficients for the indicator variables from D92 to D96 are all negative, and there is a tendency to become more and more negative. If fixed the size and age of the house, the house prices stable negative growth.

(c) What would have happened if we had included an indicator variable for 1991?

The equation's years are from D92 to D96, if we add the indicator variable for 1991 will change the equation as below:

δ_1D92+ δ_2D93 + δ_3D94 + δ_4D95 + δ_5D96 +δ_6D91 will equal to one.

The above equation is failing to omit one indicator variable, which is D91 for EXERCISE 7.4. This leads to exact multi-collinearity, and exact collinearity would cause the least-squares estimation to fail.

EXERCISE 7.15:

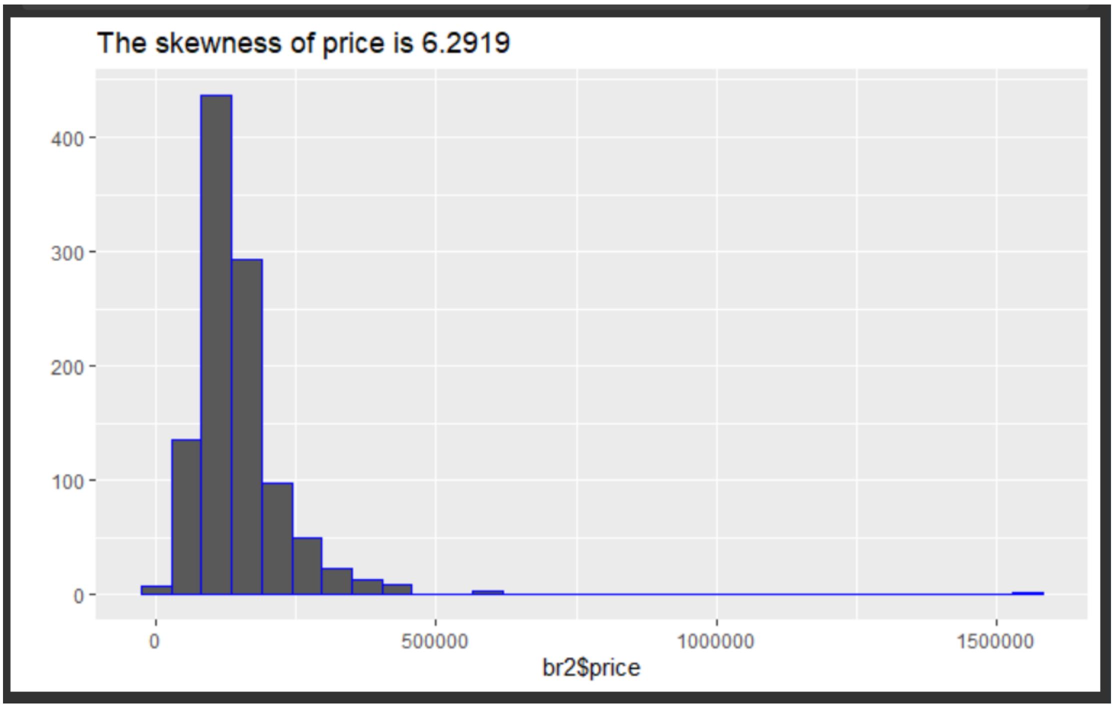
The data file *br2.dat* contains data on 1080 house sales in Baton Rouge, Louisiana, during July and August 2005. The variables are PRICE ($), SQFT (total square feet), BEDROOMS (number), BATHS (number), AGE (years), OWNER (= 1 if occupied by owner; zero if vacant or rented), POOL (= 1 if present), TRADITIONAL (= 1 if traditional style; 0 if other style), FIREPLACE (= 1 if present), and WATERFRONT (= 1 if on waterfront).
(a) Compute the data summary statistics and comment. In particular, construct a histogram of PRICE. What do you observe?

```
```{r}
# Check the br2.dat dataset.
library(foreign)
br2 <- read.dta("br2.dta")
summary(br2)
```
```

```
     price              sqft         bedrooms         baths             age
 Min.   :  22000   Min.   : 662   Min.   :1.00   Min.   :1.000   Min.   : 1.00
 1st Qu.:  99000   1st Qu.:1604   1st Qu.:3.00   1st Qu.:2.000   1st Qu.: 5.00
 Median : 130000   Median :2186   Median :3.00   Median :2.000   Median :18.00
 Mean   : 154863   Mean   :2326   Mean   :3.18   Mean   :1.973   Mean   :19.57
 3rd Qu.: 170163   3rd Qu.:2800   3rd Qu.:4.00   3rd Qu.:2.000   3rd Qu.:25.00
 Max.   :1580000   Max.   :7897   Max.   :8.00   Max.   :5.000   Max.   :80.00
     owner             pool          traditional       fireplace       waterfront
 Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.000   Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.00000
 Median :0.0000   Median :0.00000   Median :1.0000   Median :1.000   Median :0.00000
 Mean   :0.4889   Mean   :0.07963   Mean   :0.5389   Mean   :0.563   Mean   :0.07222
 3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.00000
 Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.000   Max.   :1.00000
```

```
```{r}
# Compute a histogram of 'br2$price'
library(ggplot2)
library(moments)
skewness(br2$price)
qplot(br2$price, geom="histogram",  main = 'The skewness of price is 6.2919' ,col=I("blue"))
```
```

Jeffrey Chang
jlc190004



- the distribution of PRICE is positively skewed.
- the median price $130,000 is very different from the maximum price of $1,580,000.

(b) Estimate a regression model explaining ln (PRICE=1000) as a function of the remaining variables. Divide the variable SQFT by 100 prior to estimation. Comment on how well the model fits the data. Discuss the signs and statistical significance of the estimated coefficients. Are the signs what you expect? Give an exact interpretation of the coefficient of WATERFRONT.

```
```{r}
mod1 <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront, data=br2)
summary(mod1)
```
```

```
Call:
lm(formula = log(price/1000) ~ sqft + bedrooms + baths + age +
    owner + pool + traditional + fireplace + waterfront, data = br2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.13459 -0.12758  0.00656  0.14785  1.06650

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.981e+00  4.589e-02  86.738  < 2e-16 ***
sqft         2.990e-04  1.406e-05  21.269  < 2e-16 ***
bedrooms    -3.151e-02  1.661e-02  -1.897 0.058135 .
baths        1.901e-01  2.056e-02   9.248  < 2e-16 ***
age         -6.215e-03  5.179e-04 -11.999  < 2e-16 ***
owner        6.747e-02  1.775e-02   3.802 0.000152 ***
pool        -4.275e-03  3.158e-02  -0.135 0.892353
traditional -5.609e-02  1.703e-02  -3.294 0.001019 **
fireplace    8.427e-02  1.901e-02   4.432 1.03e-05 ***
waterfront   1.100e-01  3.336e-02   3.297 0.001010 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.27 on 1070 degrees of freedom
Multiple R-squared:  0.7373,    Adjusted R-squared:  0.7351
F-statistic: 333.7 on 9 and 1070 DF,  p-value: < 2.2e-16
```

- The estimated model fits the data with R^2 = 0.737 and adjusted R^2= 0.7351.
- SQFT: The estimated coefficient is positive and significant, indicating that an extra 100 ft square of living space, other variables fixed, will make the house price increase approximately 3%.
- BEDROOMS: The estimated coefficient is negative and significant at 0.1 level, indicating that an extra bedroom, other variables fixed, will make the house price decrease approximately 3.15%.
- BATHS: The estimated coefficient is positive and significant, indicating that an extra bath, other variables fixed, will make the house price increase approximately 19%.
- AGE: Depreciation reduces the value of the home by 0.62 % per year
- OWNER: Homes with owner live-in are estimated to sell for 6.7% more than empty houses. It is positive and significant.
- POOL: The estimated coefficient is negative and statistically insignificant. The pool will somehow decrease the house price.
- TRADITIONAL: this style of house will sell 5.6% less price.
- FIREPLACE: It is positive and significant estimated coefficient. Have fireplace estimated 8.4% increase in the house value.
- WATERFRONT: A waterfront house sells for 11.62% higher than a house without waterfront. 100($e^{(0.1100)}$ – 1) = 11.62%

(c) Create a variable that is the product of WATERFRONT and TRADITIONAL. Add this variable to the model and re-estimate. What is the effect of adding this variable? Interpret the coefficient of this interaction variable and discuss its sign and statistical significance.

<u>Previous model Table:</u>

```
```{r}
mod1 <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront, data=br2)
kable(tidy(mod1), caption="A Regression Model", digits=4)
```
```

term	estimate	std.error	statistic	p.value
(Intercept)	3.9808	0.0459	86.7384	0.0000
sqft	0.0003	0.0000	21.2686	0.0000
bedrooms	-0.0315	0.0166	-1.8967	0.0581
baths	0.1901	0.0206	9.2480	0.0000
age	-0.0062	0.0005	-11.9985	0.0000
owner	0.0675	0.0177	3.8017	0.0002
pool	-0.0043	0.0316	-0.1354	0.8924
traditional	-0.0561	0.0170	-3.2944	0.0010
fireplace	0.0843	0.0190	4.4320	0.0000
waterfront	0.1100	0.0334	3.2970	0.0010

The following is new model Table:

```r
# create new model table (traditional*waterfront)
mod2 <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront+traditional*waterfront,
data=br2)
kable(tidy(mod2), caption="A Regression Model_2 ", digits=4)
```

term	estimate	std.error	statistic	p.value
(Intercept)	3.9711	0.0459	86.4301	0.0000
sqft	0.0003	0.0000	21.3989	0.0000
bedrooms	-0.0313	0.0166	-1.8909	0.0589
baths	0.1883	0.0205	9.1740	0.0000
age	-0.0061	0.0005	-11.8811	0.0000
owner	0.0684	0.0177	3.8614	0.0001
pool	-0.0024	0.0315	-0.0760	0.9395
traditional	-0.0449	0.0176	-2.5575	0.0107
fireplace	0.0873	0.0190	4.5938	0.0000
waterfront	0.1654	0.0400	4.1395	0.0000
traditional:waterfront	-0.1722	0.0687	-2.5056	0.0124

- The approximate percentage difference in price on no waterfront with traditional house is -4.49%. The exact percentage price difference is 100(e^{δ}-1) %=100($e^{-0.0449}$-1) % = -4.39%.
- The approximate percentage difference in price on traditional house with waterfront is (-0.0449+0.1654-0.1722) = 5.17%. The approximate percentage difference is 100(e^{δ}-1) %=100($e^{(0.0517)}$-1) % = -5.04%
- The approximate percentage difference in price on nontraditional house with waterfront is 16.54%. The exact percentage price difference is 100(e^{δ}-1) % = 100($e^{0.1654}$-1) % = 17.99%.
- The traditional houses on the waterfront sell for less than traditional houses elsewhere. (-5.04% < -4.39%)
- The price advantage from being on the waterfront is lost if the house is a traditional style.
- The extra effect from both characteristics, (Traditional × Waterfront), must also be added. Its estimate is significant at a 5% level of significance, (p-value = 0.0124).
- we need to calculate γ for those houses which are traditional style and on the waterfront.

Jeffrey Chang
jlc190004

(d) It is arguable that the traditional-style homes may have a different regression function from the diverse set of nontraditional styles. Carry out a Chow test of the equivalence of the regression models for traditional versus nontraditional styles. What do you conclude?

```
```{r}
mod3 <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront, data=br2) # no traditional variable

dnotrad <- br2[which(br2$traditional==0),]
dtrad <- br2[which(br2$traditional==1),]

mod5not <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront+traditional*waterfront,
data=dnotrad)# traditional=0

mod5t <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront+traditional*waterfront,
data=dtrad)#traditional=1

mod6 <-lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront+traditional*waterfront+traditional/
            (sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront+traditional*waterfront), data=br2)


stargazer(mod3, mod5t, mod5not, mod6, header=FALSE,
type='text',
title="Model comparison, 'Price' equation",
keep.stat="n",digits=2, single.row=TRUE,
intercept.bottom=FALSE)
```
```

Model comparison, 'Price' equation

	Dependent variable:			
	log(price/1000)			
	(1)	(2)	(3)	(4)
Constant	3.97*** (0.05)	3.73*** (0.07)	4.07*** (0.07)	4.07*** (0.06)
sqft	0.0003*** (0.0000)	0.0003*** (0.0000)	0.0003*** (0.0000)	0.0003*** (0.0000)
bedrooms	-0.04** (0.02)	0.03 (0.02)	-0.07*** (0.03)	-0.07*** (0.02)
baths	0.19*** (0.02)	0.21*** (0.03)	0.18*** (0.03)	0.18*** (0.03)
age	-0.01*** (0.001)	-0.01*** (0.001)	-0.01*** (0.001)	-0.01*** (0.001)
owner	0.07*** (0.02)	0.10*** (0.02)	0.04 (0.03)	0.04 (0.03)
pool	0.001 (0.03)	-0.02 (0.04)	0.002 (0.05)	0.002 (0.04)
traditional				-0.34*** (0.09)
waterfront:traditional				-0.21*** (0.07)
sqft:traditional				-0.0001* (0.0000)
bedrooms:traditional				0.10*** (0.03)
baths:traditional				0.03 (0.04)
age:traditional				-0.001 (0.001)
owner:traditional				0.06* (0.04)
pool:traditional				-0.02 (0.06)
fireplace:traditional				0.07* (0.04)
fireplace	0.09*** (0.02)	0.12*** (0.02)	0.06* (0.03)	0.06* (0.03)
waterfront	0.12*** (0.03)	-0.03 (0.05)	0.17*** (0.05)	0.17*** (0.04)
traditional:waterfront				
Observations	1,080	582	498	1,080
Note:				*p<0.1; **p<0.05; ***p<0.01

The above have (1)~(4) model.

The restricted model (1) is assumed that there is no difference between TRADITIONAL and non-traditional houses (Rest).

Two models are for the subsets of the data for which the variable TRADITIONAL = 1 (2) or TRADITIONAL = 0 (3)

The last model (4) is the fully interacted model.

The *F*-value for this test is:

```
```{r}
mod3 <- lm(log(price/1000)~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront, data=br2)
kable(anova(mod3, mod6),
caption="Chow test for the 'Price' equation")
```
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1071	78.77189	NA	NA	NA	NA
1062	75.79949	9	2.972398	4.627247	5e-06

((78.77189-75.79949)/9)/ (75.79949/ (1080-18)) = 4.62725

Since 4.62725 > 1.889 = F_(0.95,9,1062), so rejected the null hypothesis at α = 0.05.

We conclude traditional style and non-traditional style regression functions have differences.

(e) Using the equation estimated in part (c)*(correct textbook error),* **predict the value of a traditional style house with 2500 square feet of area, that is 20 years old, that is owner-occupied at the time of sale, that has a fireplace, 3 bedrooms, and 2 baths, but no pool, and that is not on the waterfront.**

Function = 3.9711+0.0003*sqft-0.0313*bedrooms+0.1883*baths-0.0061*age+0.0684*owner-0.0024*pool-0.0449*traditional+0.0873*fireplace+0.1654*waterfront-0.1722*(trad*water)

Function(value-in) = 3.9711+0.0003*2500-0.0313*3+0.1883*2-0.0061*20+0.0684*1-0.0024*0-0.0449*1+0.0873*1+0.1654*0-0.1722*(1*0) =4.9926

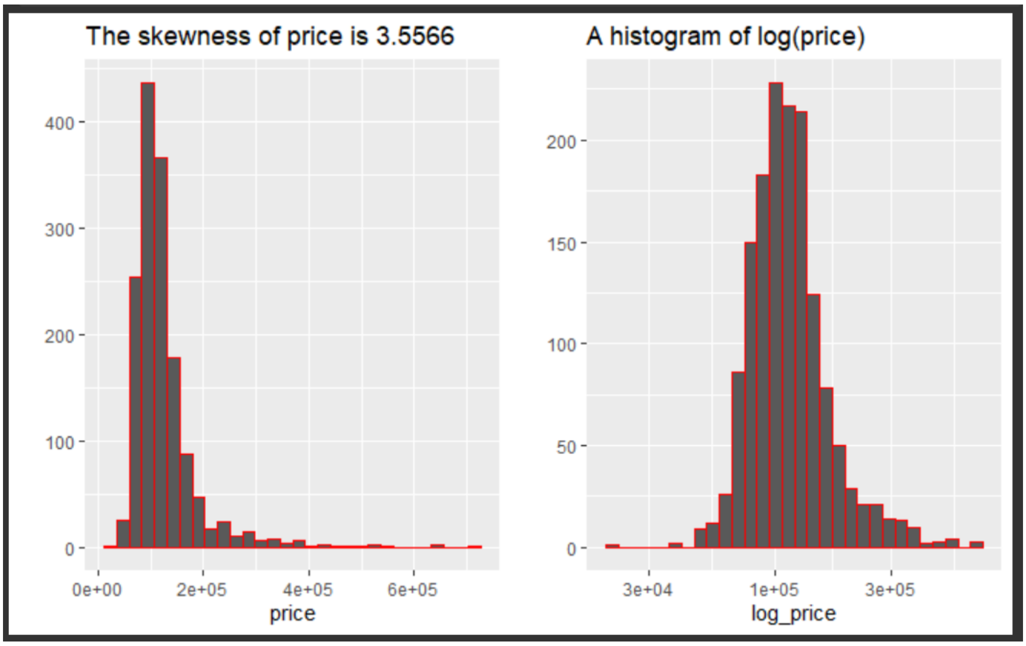
So, the estimated predict value = $\widehat{PRICE}$ = $e^{(4.9926)}$*1000 = $147,319

<u>EXERCISE 7.16:</u>

Data on 1500 house sales from Stockton, California, are contained in the data file *<u>stockton4.dat</u>*. [Note: stockton3.dat is a larger version of the same data set, containing 2610 observations.] The houses are detached single-family homes that were listed for sale between October 1, 1996, and November 30, 1998. The variables are PRICE ($), LIVAREA (hundreds of square feet), BEDS (number of bedrooms), BATHS (number of bathrooms), LGELOT (= 1 if lot size is greater than 0.5 acres, zero otherwise), AGE (years), and POOL (= 1 if home has pool, zero otherwise).

(a) Examine the histogram of PRICE. What do you observe? Create the variable ln(PRICE) and examine its histogram. Comment on the difference.

````
```{r}
stt4 <- read.dta("stockton4.dta")
library(gridExtra)
price <- stt4$sprice
log_price <- stt4$sprice
skewness(stt4$sprice)
plot1 <- qplot(price, geom="histogram",  main = 'The skewness of price is 3.5566' ,col=I("red"))
plot2 <- qplot(log_price, log = 'x', geom="histogram",  main = 'A histogram of log(price)' ,col=I("red"))
grid.arrange(plot1, plot2, ncol=2)
```
````



The histogram for *PRICE* is positively skewed. The ln (*PRICE*) is less skewed and is more like symmetrical. Thus, the histogram of the ln (*PRICE*) is closer in shape to a normal distribution than the histogram of *PRICE*.

(b) Estimate a regression of ln(PRICE/1000) on the remaining variables. Discuss the estimation results. Comment on the signs and significance of the variables LIVAREA, BEDS, BATHS, AGE, and POOL.

````
```{r}
mod7 <- lm(log(price/1000)~livarea+beds+baths+lgelot+age+pool, data=stt4)
kable(tidy(mod7), caption="A Regression Model_7.16(b) ", digits=4)
```
````

term	estimate	std.error	statistic	p.value
(Intercept)	3.9860	0.0373	106.7462	0.0000
livarea	0.0539	0.0017	31.5764	0.0000
beds	-0.0382	0.0114	-3.3647	0.0008
baths	-0.0103	0.0165	-0.6216	0.5343
lgelot	0.2531	0.0255	9.9103	0.0000
age	-0.0013	0.0005	-2.8500	0.0044
pool	0.0787	0.0231	3.4119	0.0007

The estimated equation is

$ln\ (\widehat{PRICE}/1000)$ = 3.9860 + 0.539*(LIVAREA) - 0.0382*(BEDS) - 0.0103*(BATHS) + 0.253*(LGELOT) - 0.0013*(AGE) + 0.0787*(POOL)

According to the p-value, all coefficients are significant except for BATHS.

LIVAREA: It is reasonable that a bigger area has a higher price when holding all else fixed.

BEDS: When holding all else fixed, more rooms mean each room is smaller, so the house price is lower.

BATHS: The number of baths is statistically insignificant, so it is hard to interpret.

AGE: It is reasonable that an older house has a lower price, when holding all else fixed.

POOL: It is reasonable that the house has a pool is expensive than no pool house, when holding all else fixed.

(c) Discuss the effect of large lot size on the selling price of a house.

LGELOT (= 1 if lot size is greater than 0.5 acres, zero otherwise).

The price of houses on lot sizes greater than 0.5 acres is approximately 100($e^{(0.2531)}$ – 1) = 28.8% larger than the price of houses on lot sizes less than 0.5 acres.

(d) Introduce to the model an interaction variable LGELOT*LIVAREA. Estimate this model and discuss the interpretation, sign, and significance of the coefficient of the interaction variable.

```{r}
mod8 <- lm(log(price/1000)~livarea+beds+baths+lgelot+age+pool+lgelot*livarea, data=stt4)
kable(tidy(mod8), caption="A Regression Model_7.16(d) ", digits=4)
```

term	estimate	std.error	statistic	p.value
(Intercept)	3.9649	0.0370	107.0645	0.0000
livarea	0.0589	0.0019	31.5824	0.0000
beds	-0.0480	0.0113	-4.2368	0.0000
baths	-0.0201	0.0164	-1.2234	0.2214
lgelot	0.6134	0.0632	9.7050	0.0000
age	-0.0016	0.0005	-3.5269	0.0004
pool	0.0853	0.0228	3.7442	0.0002
livarea:lgelot	-0.0161	0.0026	-6.2174	0.0000

$ln\ (\widehat{PRICE}/1000)$ = 3.9649+0.0589*(LIVAREA)-0.0480*(BEDS)-0.0201*(BATHS)+0.6134*(LGELOT)-0.0016*(AGE)+0.0853*(POOL)-0.0161*(LGELOT ×LIVAREA)

Interpretation of the coefficient of LGELOT × LIVAREA:

The estimated marginal effect of a 100 sq ft gain in living area in a house on a lot of less than 0.5 acres is 5.89 percent, keeping other factors unchanged.

Jeffrey Chang
jlc190004

This is estimated if the same rise for a house on a large lot raises the selling price of the property by 1.61% less, or 4.27%. the LGELOT coefficient improves significantly.

(e) Carry out a Chow test of the equivalence of models for houses that are on large lots and houses that are not.

```r
mod9 <- lm(log(sprice/1000)~livarea+beds+baths+age+pool, data=stt4) # no lot

dnolot <- stt4[which(stt4$lgelot==0),]
dlot <- stt4[which(stt4$lgelot==1),]

mod10n <- lm(log(sprice/1000)~ livarea+beds+baths+lgelot+age+pool+(lgelot*livarea), data=dnolot)# lot=0

mod10<-lm(log(sprice/1000)~livarea+beds+baths+lgelot+age+pool+(lgelot*livarea), data=dlot) #lot=1

mod11 <- lm(log(sprice/1000)~livarea+beds+baths+age+pool+lgelot*livarea+lgelot/(livarea+beds+baths+age+pool+lgelot*livarea)
        ,data = stt4)

stargazer(mod9, mod10, mod10n, mod11, header=FALSE,
type='text',
title="Model comparison, 'sprice' equation",
keep.stat="n",digits=2, single.row=TRUE,
intercept.bottom=FALSE)
```

Model comparison, 'sprice' equation

	Dependent variable: log(sprice/1000)			
	(1)	(2)	(3)	(4)
Constant	3.98*** (0.04)	4.41*** (0.18)	3.98*** (0.04)	3.98*** (0.04)
livarea	0.06*** (0.002)	0.03*** (0.01)	0.06*** (0.002)	0.06*** (0.002)
beds	-0.06*** (0.01)	-0.01 (0.05)	-0.05*** (0.01)	-0.05*** (0.01)
baths	-0.03 (0.02)	0.08 (0.07)	-0.03** (0.02)	-0.03* (0.02)
lgelot				0.43*** (0.14)
age	-0.001* (0.0005)	-0.002 (0.002)	-0.002*** (0.0005)	-0.002*** (0.0005)
pool	0.10*** (0.02)	0.13* (0.07)	0.07*** (0.02)	0.07*** (0.03)
livarea:lgelot				-0.03*** (0.004)
beds:lgelot				0.04 (0.04)
baths:lgelot				0.12** (0.05)
age:lgelot				-0.0002 (0.001)
pool:lgelot				0.06 (0.06)
Observations	1,500	95	1,405	1,500
Note:				*p<0.1; **p<0.05; ***p<0.01

```r
kable(anova(mod9, mod11),
caption="Chow test for the 'sPrice' equation")
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1494	72.06331	NA	NA	NA	NA
1488	65.47123	6	6.592085	24.97031	0

The value of the F-statistic is:

((72.06331-65.47123)/6)/ (65.4712/ (1488)) = 24.97

Since 24.97 > 2.10 = F_(0.95,6,1488), so rejected the null hypothesis at α = 0.05.

We conclude the pricing structure for houses on large lots is not the same as that on smaller lots.