MSIT 431: PROBABILITY AND STATISTICAL METHODS

Submitted By : Ruma Anand

Student ID : 3074142

Ques 1. Textbook Exercise 8.63

Solution 1.

A) Explanatory Variable: Type of College as Public or private distinction can be used to explain the observed outcomes.
   Response Variable: Whether physical education is required or not.
B) Populations are the public and private colleges and universities.
C) Below mentioned are the statistics:-
   X1=101, n1=129, p1 hat =101/129 =0.7829
   X2=60, n2=225, p2 hat=60/225 = 0.2667
D) Standard Error= $((p1(1-p1))/n1 + (p2(1-p2))/n2)^{1/2}$
   Standard Error = $(0.1699/129+0.1955/225)^{1/2}$
   $=(0.0013+0.00087)^{1/2}$
   =0.046
   For confidence interval 95%, Z*=1.96

   M=Z*xSE=1.96*0.046 =0.09016

   (P1-p2)±M=0.5162±0.09016

   =(0.4260,0.6079)

E) H0 : p1=p2
   Ha:p1≠p2
   P=(60+101)/(225+129)
   =0.4548

   $Z=(p1-p2)/((p(1-p)(1/n1+1/n2))^{1/2}$

   =-9.387
   P value=0

F) The number of successes and failures are greater than 5,but unsure if samples were selected randomly.

G) We can conclude that physical education requirement in public colleges is higher than private colleges.

Ques 2. Textbook Exercise 8.75

Solution 2.

A) Sample Size : 2432

Count : 1639

B) p=1639/2342 =0.6998

Standard Error=$(p(1-p)/n)^{1/2}$

$=(0.6998(0.3002)/2432)^{1/2}$
$=0.0092$

C)p±z*SE
=0.6998±0.0092
=(0.6817,0.718)

D)Guidelines for the use of large sample confidence interval is satisfied as number of successes and failures is greater than 15.

Ques 3. Textbook Exercise 8.77

Solution 3.

H1=1055

P1=861/(861+194)

=.8161

H2=974

P2=417/(417+557)

=0.4281

Standard error = $(p1(1-p1)/n1 + p2(1-p2)/n2)^{1/2}$

=.01984

P1-p2±1.96*.01984

=(.3491,.4269)

We can use the large-sample procedures for this comparison because the counts of "successes" and "failures" in each group are all at least 10.

Ques 4. Textbook Exercise 8.95

Solution 4.

a) C= 95%,

z^*=1.96

m≤0.055,

$\hat{p}$=0.5

m=0.055=1.96* $\sqrt{\frac{0.5*0.5}{n}+\frac{0.5*0.5}{n}}$, so n=634.975

b) n= $(\frac{z*}{m})^2$*0.5

Ques 5. Textbook Exercise 10.42

Solution 5.

a) β_0=-0.08, it is the intercept When U.S.RETURN=0, overseas =-0.08

b) β_1=0.20, it is the slope. It shows that the relationship between U.S. and overseas returns is positive and linear.

c) The regression model is y_i=-0.08+0.20*x_i+e_i.

-0.08+0.20*x_i is the mean response when x=x_i.

e_i= residual

Ques 6. Textbook Exercise 10.45

Solution 6.

a) 30 homes have sales price greater than the assessed value.

Let the null hypothesis be H_0: $\mu_{salesprice-assessedvalue}$=0

Assumption : H_a: $\mu_{salesprice-assessedvalue}$ >0

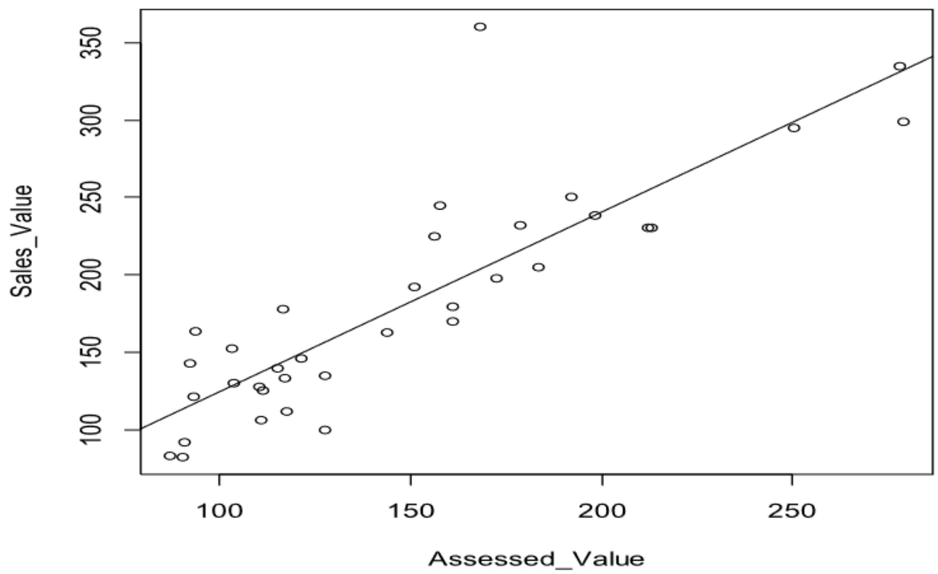
μ_{diff}=32.28, S_{diff}=37.7075

$$t=\frac{32.28}{37.7075/\sqrt{35}}=5.0645$$

The p value will be very small, rejecting the null hypothesis. the trend would be true for the larger population of all homes recently sold.

b) The relationship is positive, linear and strong.

> plot(Assessed_Value,Sales_Price)



c) Property 27 is an unusual observation. This can influence the least-squares line. The slope decreases if property 27 is removed.

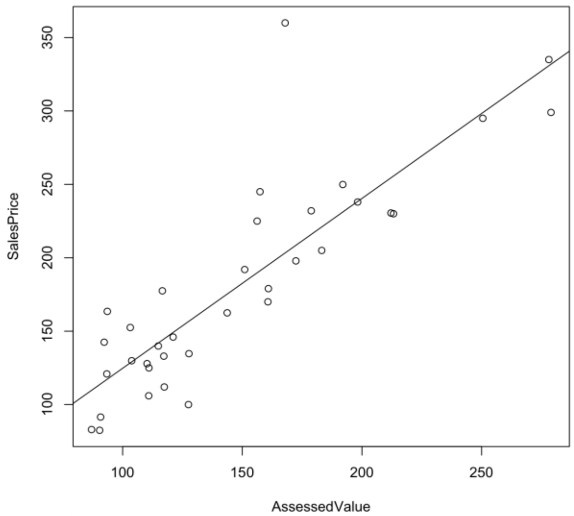
d) > modelAS<-lm(SalesPrice~AssessedValue)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.0176	19.1044	0.472	0.64	
AssessedValue	1.1570	0.1217	9.505	5.69e-11	***

Residual standard error: 37.34 on 33 degrees of freedom, so the estimated model standard error is 37.34.

The least-squares regression line is y=9.0176+1.157x.



e) > Filtereddata<-SALES[-27,]

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.43181	12.98994	0.726	0.473	
Filtereddata$AssessedValue	1.12300	0.08295	13.539	8.56e-15	***

Residual standard error: 25.39

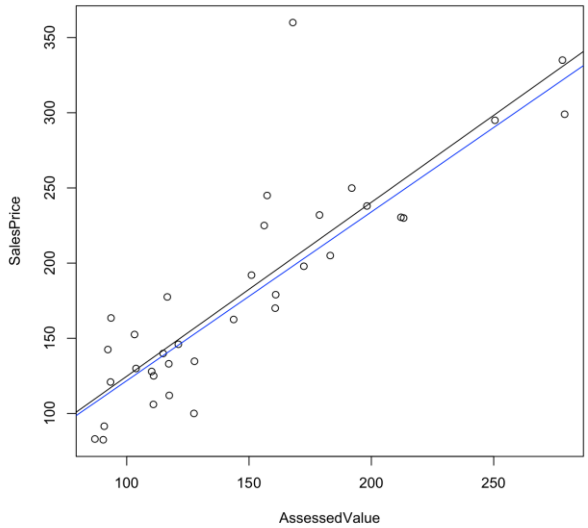
Multiple R-squared: 0.8514, Adjusted R-squared: 0.8467

F-statistic: 183.3 on 1 and 32 DF,

p-value: 8.563e-15

SE = 25.39

the least-squares regression line is y=9.43181+1.123x.

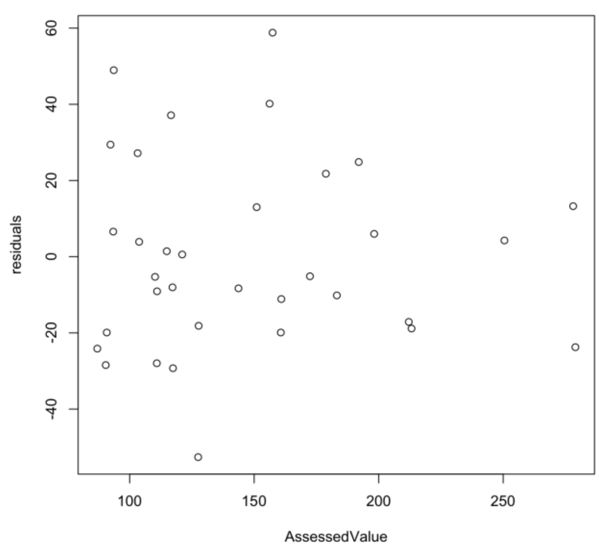


f) Based on analysis in part d and e, we know that the estimated model standard error is smaller when home 27 is not considered.

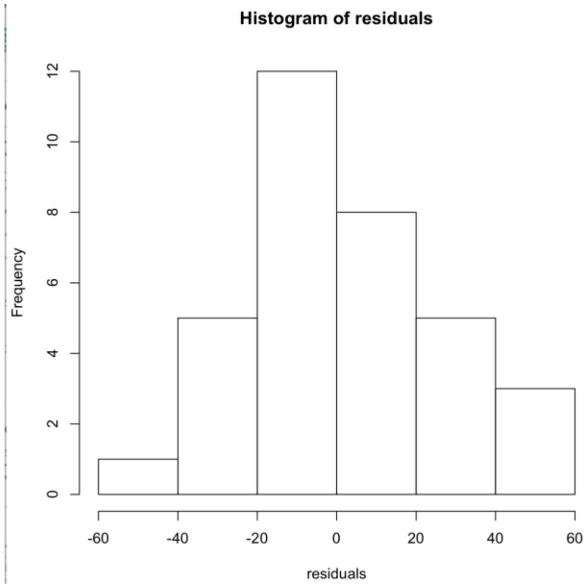
g) It is more appropriate to just consider the 34 properties, because the standard error is smaller, which means the predicted values are more accurate.

Ques 7. Textbook Exercise 10.46

a) Below is a scatterplot of residuals and assessed value. There is nothing unusual to report.



b) Below is a histogram of residuals. We can see it’s approximately Normal.



c) Based on analysis in part a and part b, the assumptions for statistical inference are reasonably satisfied as and data points appear to be normally distributed.

d) C=95%, df=n-2=32,

SE_{b1}= 0.08295, the mean of b_1=1.123for df = 32 at 95% confidence = 2.0345

$b_1 \pm$t*SE_{b1} =

=1.123 ± 2.0345 * 0.08295

=1.123 ± 0.1688

confidence interval for the slope is:

= (0.9542, 1.2918)

e) When y=x, the slope is 1, we can do a significance test.

H_0: β_1=1

H_a: β_1 ≠1

$> \ t = (1.123 - 1)/0.08295$

$> \ pt(t, df\ = 32, lower.tail = F)$

[1] 0.07395227 > 0.05

Therefore, we do not reject H_0, which means there is no evidence that this model is not reasonable. The selling price can be larger or smaller or equal to the assessed value.