**Question 1:**

(a)

*CODE:*

>data fall;

> infile '/home/shu.tu/fall.txt';

> input time height;

>

>proc gplot;

> plot height\*time;

*RESULTS:*



(b)

From the plot we can see that: i)direction: approximately negative gradient, as "time"(explanatory variable) increases, "height"(response variable) decreases, except in the beginning(from 0 to 1), it goes up; ii)form: it is a curve, so there is no liner association between "time" and "height"; iii)strength: all the points are on the curve, therefore there is a strong negative correlation.

(c)

*CODE:*

>proc reg;

> model height=time;

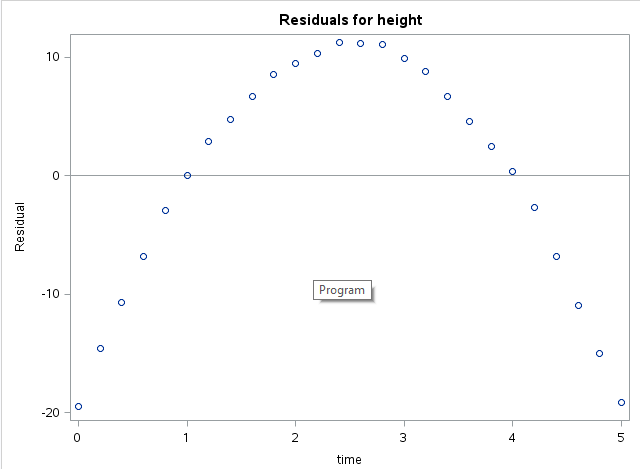
> output out=util2 r=res p=fit;

>

>proc gplot;

> plot res\*fit;

*RESULTS:*



Apparently, a straight line fits badly.

(d)

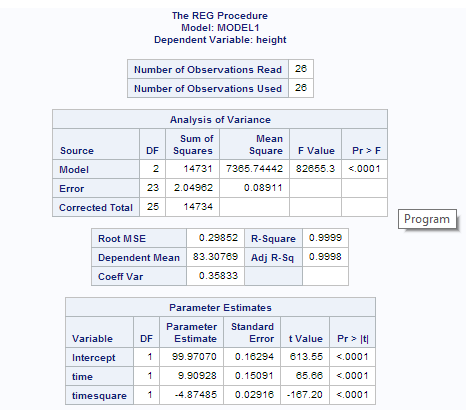
*CODE:*

> timesquare=time\*time;

>proc reg;

> model height=time timesquare;

*RESULTS:*



(e)

The p-value for "time square" is less than 0.001 which is very small, therefore we conclude that it is significant. It does not surprise me, since "time" and "height" have a strong correlation, which implies that "time" has strong explanatory power for "height" and then the square of "time" must have even more explanatory power for "height" (i.e., significant).

(f)

From the table we know that: height = 99.97070 + 9.90928t - 4.87485t2, compare to the formula s=ut+0.5\*at2

we have: the height of the observation deck of the tower is around 100, the velocity upwards that the ball thrown was 9.90928 and the acceleration here is -4.87485/0.5=-9.4797.

**Question 2:**

(a)

> mesq=read.table("mesquite.txt",header=T)

> head(mesq)

Obs Group Diam1 Diam2 TotHt CanHt Dens LeafWt

1 1 MCD 1.8 1.15 1.30 1.00 1 401.3

2 2 MCD 1.7 1.35 1.35 1.33 1 513.7

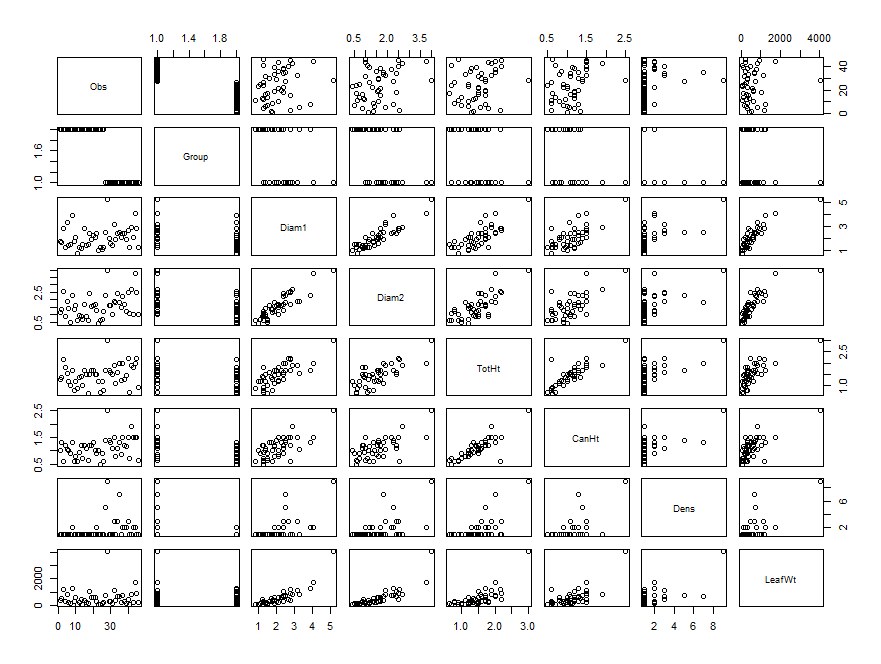
3 3 MCD 2.8 2.55 2.16 0.60 1 1179.2

4 4 MCD 1.3 0.85 1.80 1.20 1 308.0

5 5 MCD 3.3 1.90 1.55 1.05 1 855.2

6 6 MCD 1.4 1.40 1.20 1.00 1 268.7

> plot(mesq)



(b)

From the pairs plot we can see that "Diam1" "Diam2" "TotHt" "CanHt" all go up with "Leafwt" (i.e., all of them are positively related to Leafwt); "TotHt" and "CanHt" are positively related with some outliers; "Diam1" and "Diam2" are positively related as well with some outliers.

(c)

> apply(mesq[3:8],2,quantile)

Diam1 Diam2 TotHt CanHt Dens LeafWt

0% 0.800 0.400 0.65 0.5000 1 60.200

25% 1.400 1.000 1.20 0.8625 1 219.625

50% 1.950 1.525 1.50 1.1000 1 361.850

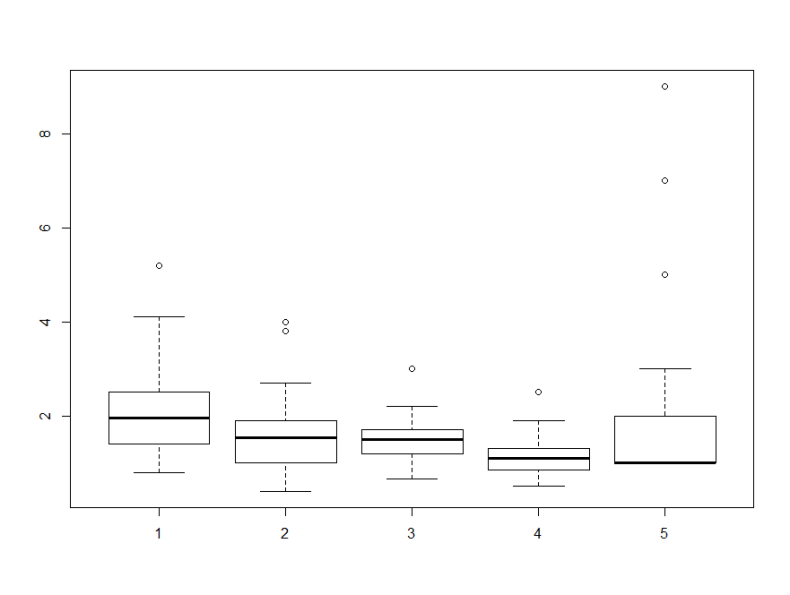
75% 2.475 1.900 1.70 1.3000 2 688.725

100% 5.200 4.000 3.00 2.5000 9 4052.000

(d)

> attach(mesq)

> boxplot(Diam1,Diam2,TotHt,CanHt,Dens)



(e)

All the variables have outlier(s).

(f)

> z=rep(1,46)

> mesq.lm=lm(z~Diam1+Diam2+TotHt+CanHt+Dens,data=mesq)

> h=hatvalues(mesq.lm)

> cb=cbind(mesq,h)

Obs Group Diam1 Diam2 TotHt CanHt Dens LeafWt h

1 1 MCD 1.80 1.15 1.30 1.00 1 401.3 0.03176992

2 2 MCD 1.70 1.35 1.35 1.33 1 513.7 0.08981562

3 3 MCD 2.80 2.55 2.16 0.60 1 1179.2 0.58981729

4 4 MCD 1.30 0.85 1.80 1.20 1 308.0 0.15904390

5 5 MCD 3.30 1.90 1.55 1.05 1 855.2 0.19761446

6 6 MCD 1.40 1.40 1.20 1.00 1 268.7 0.07002485

7 7 MCD 1.50 0.50 1.00 0.90 1 155.5 0.09897821

8 8 MCD 3.90 2.30 1.70 1.30 2 1253.2 0.25460805

9 9 MCD 1.80 1.35 0.80 0.60 1 328.0 0.11056986

10 10 MCD 2.10 1.60 1.20 0.80 1 614.6 0.05330413

11 11 MCD 0.80 0.63 0.90 0.60 1 60.2 0.09125336

12 12 MCD 1.30 0.95 1.35 0.95 1 269.6 0.05013489

13 13 MCD 1.20 0.90 1.40 1.20 1 448.4 0.08285226

14 14 MCD 1.50 0.70 1.00 0.70 1 120.4 0.06811367

15 15 MCD 2.80 1.70 1.70 1.20 1 378.7 0.10131484

16 16 MCD 1.40 0.85 1.50 1.10 1 266.4 0.07118072

17 17 MCD 1.50 0.60 0.65 0.64 1 138.9 0.13349732

18 18 MCD 2.40 2.40 1.50 1.20 1 1020.8 0.11051371

19 19 MCD 1.90 1.55 1.70 1.20 1 635.7 0.05291006

20 20 MCD 2.30 1.60 1.70 1.30 1 621.8 0.05395073

21 21 MCD 2.10 1.70 1.50 1.00 1 579.8 0.03648040

22 22 MCD 2.40 1.30 1.50 0.90 2 326.8 0.10269126

23 23 MCD 1.00 0.40 1.20 1.00 1 66.7 0.09087520

24 24 MCD 1.30 0.60 0.70 0.50 1 68.0 0.10744979

25 25 MCD 1.10 0.70 1.20 0.90 1 153.1 0.05706602

26 26 MCD 1.30 1.20 0.80 0.60 1 256.4 0.10795370

27 27 ALS 2.50 2.30 1.70 1.40 5 723.0 0.18243805

28 28 ALS 5.20 4.00 3.00 2.50 9 4052.0 0.53643328

29 29 ALS 2.00 1.60 1.70 1.40 1 345.0 0.06768650

30 30 ALS 1.60 1.60 1.60 1.30 1 330.9 0.08860010

31 31 ALS 1.40 1.00 1.50 1.10 1 163.5 0.06066513

32 32 ALS 3.20 1.90 1.90 1.50 3 1160.0 0.11612343

33 33 ALS 1.90 1.80 1.10 0.80 1 386.6 0.09019280

34 34 ALS 2.40 2.40 1.60 1.10 3 693.5 0.11636855

35 35 ALS 2.50 1.80 2.00 1.30 7 674.4 0.40084821

36 36 ALS 2.10 1.50 1.25 0.85 1 217.5 0.04294228

37 37 ALS 2.40 2.20 2.00 1.50 2 771.3 0.07712944

38 38 ALS 2.40 1.70 1.30 1.20 2 341.7 0.06515208

39 39 ALS 1.90 1.20 1.45 1.15 2 125.7 0.03717438

40 40 ALS 2.70 2.50 2.20 1.50 3 462.5 0.10350085

41 41 ALS 1.30 1.10 0.70 0.70 1 64.5 0.12190012

42 42 ALS 2.90 2.70 1.90 1.90 1 850.6 0.28241422

43 43 ALS 2.10 1.00 1.80 1.50 2 226.0 0.15130999

44 44 ALS 4.10 3.80 2.00 1.50 2 1745.1 0.27618888

45 45 ALS 2.80 2.50 2.20 1.50 1 908.0 0.13848106

46 46 ALS 1.27 1.00 0.92 0.62 1 213.5 0.07066646

> n=46

> k=6

> l=2\*(k+1)/n

> l

[1] 0.3043478

> cb[cb$h>=l,]

Obs Group Diam1 Diam2 TotHt CanHt Dens LeafWt h

3 3 MCD 2.8 2.55 2.16 0.6 1 1179.2 0.5898173

28 28 ALS 5.2 4.00 3.00 2.5 9 4052.0 0.5364333

35 35 ALS 2.5 1.80 2.00 1.3 7 674.4 0.4008482

The 3rd, 28th and 35th observations are outliers. There is no evidence supporting they are outliers which makes them really unusual.

(g)

Yes, it would be a good idea to transform some or all of the variables, because there are some outliers in the model and we cannot decide which of the explanatory variables are important for response variable by current variables.

(h)

> myfun=function(x) log(x)

> log.mesq=apply(mesq[3:8],2,myfun)

Diam1 Diam2 TotHt CanHt Dens LeafWt

[1,] 0.58778666 0.13976194 0.26236426 0.00000000 0.0000000 5.994709

[2,] 0.53062825 0.30010459 0.30010459 0.28517894 0.0000000 6.241639

[3,] 1.02961942 0.93609336 0.77010822 -0.51082562 0.0000000 7.072592

[4,] 0.26236426 -0.16251893 0.58778666 0.18232156 0.0000000 5.730100

[5,] 1.19392247 0.64185389 0.43825493 0.04879016 0.0000000 6.751335

[6,] 0.33647224 0.33647224 0.18232156 0.00000000 0.0000000 5.593596

[7,] 0.40546511 -0.69314718 0.00000000 -0.10536052 0.0000000 5.046646

[8,] 1.36097655 0.83290912 0.53062825 0.26236426 0.6931472 7.133456

[9,] 0.58778666 0.30010459 -0.22314355 -0.51082562 0.0000000 5.793014

[10,] 0.74193734 0.47000363 0.18232156 -0.22314355 0.0000000 6.420972

[11,] -0.22314355 -0.46203546 -0.10536052 -0.51082562 0.0000000 4.097672

[12,] 0.26236426 -0.05129329 0.30010459 -0.05129329 0.0000000 5.596939

[13,] 0.18232156 -0.10536052 0.33647224 0.18232156 0.0000000 6.105686

[14,] 0.40546511 -0.35667494 0.00000000 -0.35667494 0.0000000 4.790820

[15,] 1.02961942 0.53062825 0.53062825 0.18232156 0.0000000 5.936744

[16,] 0.33647224 -0.16251893 0.40546511 0.09531018 0.0000000 5.584999

[17,] 0.40546511 -0.51082562 -0.43078292 -0.44628710 0.0000000 4.933754

[18,] 0.87546874 0.87546874 0.40546511 0.18232156 0.0000000 6.928342

[19,] 0.64185389 0.43825493 0.53062825 0.18232156 0.0000000 6.454727

[20,] 0.83290912 0.47000363 0.53062825 0.26236426 0.0000000 6.432618

[21,] 0.74193734 0.53062825 0.40546511 0.00000000 0.0000000 6.362683

[22,] 0.87546874 0.26236426 0.40546511 -0.10536052 0.6931472 5.789348

[23,] 0.00000000 -0.91629073 0.18232156 0.00000000 0.0000000 4.200205

[24,] 0.26236426 -0.51082562 -0.35667494 -0.69314718 0.0000000 4.219508

[25,] 0.09531018 -0.35667494 0.18232156 -0.10536052 0.0000000 5.031091

[26,] 0.26236426 0.18232156 -0.22314355 -0.51082562 0.0000000 5.546739

[27,] 0.91629073 0.83290912 0.53062825 0.33647224 1.6094379 6.583409

[28,] 1.64865863 1.38629436 1.09861229 0.91629073 2.1972246 8.306966

[29,] 0.69314718 0.47000363 0.53062825 0.33647224 0.0000000 5.843544

[30,] 0.47000363 0.47000363 0.47000363 0.26236426 0.0000000 5.801816

[31,] 0.33647224 0.00000000 0.40546511 0.09531018 0.0000000 5.096813

[32,] 1.16315081 0.64185389 0.64185389 0.40546511 1.0986123 7.056175

[33,] 0.64185389 0.58778666 0.09531018 -0.22314355 0.0000000 5.957391

[34,] 0.87546874 0.87546874 0.47000363 0.09531018 1.0986123 6.541751

[35,] 0.91629073 0.58778666 0.69314718 0.26236426 1.9459101 6.513823

[36,] 0.74193734 0.40546511 0.22314355 -0.16251893 0.0000000 5.382199

[37,] 0.87546874 0.78845736 0.69314718 0.40546511 0.6931472 6.648077

[38,] 0.87546874 0.53062825 0.26236426 0.18232156 0.6931472 5.833933

[39,] 0.64185389 0.18232156 0.37156356 0.13976194 0.6931472 4.833898

[40,] 0.99325177 0.91629073 0.78845736 0.40546511 1.0986123 6.136647

[41,] 0.26236426 0.09531018 -0.35667494 -0.35667494 0.0000000 4.166665

[42,] 1.06471074 0.99325177 0.64185389 0.64185389 0.0000000 6.745942

[43,] 0.74193734 0.00000000 0.58778666 0.40546511 0.6931472 5.420535

[44,] 1.41098697 1.33500107 0.69314718 0.40546511 0.6931472 7.464567

[45,] 1.02961942 0.91629073 0.78845736 0.40546511 0.0000000 6.811244

[46,] 0.23901690 0.00000000 -0.08338161 -0.47803580 0.0000000 5.363637

> colnames(log.mesq)=c("log.Diam1","log.Diam2","log.TotHt","log.CanHt","log.Dens","log.LeafWt")

> log.mesq.pred=data.frame(Group,log.mesq)

> attach(log.mesq.pred)

> log.mesq.lm=lm(log.LeafWt~Group+log.Diam1+log.Diam2+log.TotHt+log.CanHt+log.Dens)

> pred.log.mesq=predict(log.mesq.lm,log.mesq.pred)

> head(pred.log.mesq)

1 2 3 4 5 6

5.847278 6.130674 6.947547 5.567517 6.751533 5.943203

> Leafwt.lm=lm(LeafWt~Group+Diam1+Diam2+TotHt+CanHt+Dens)

> plot(mesq[,8],mesq[,3],type="n", xlim=range(LeafWt),ylim=c(0,10),xlab="Leafwt",ylab="circum")

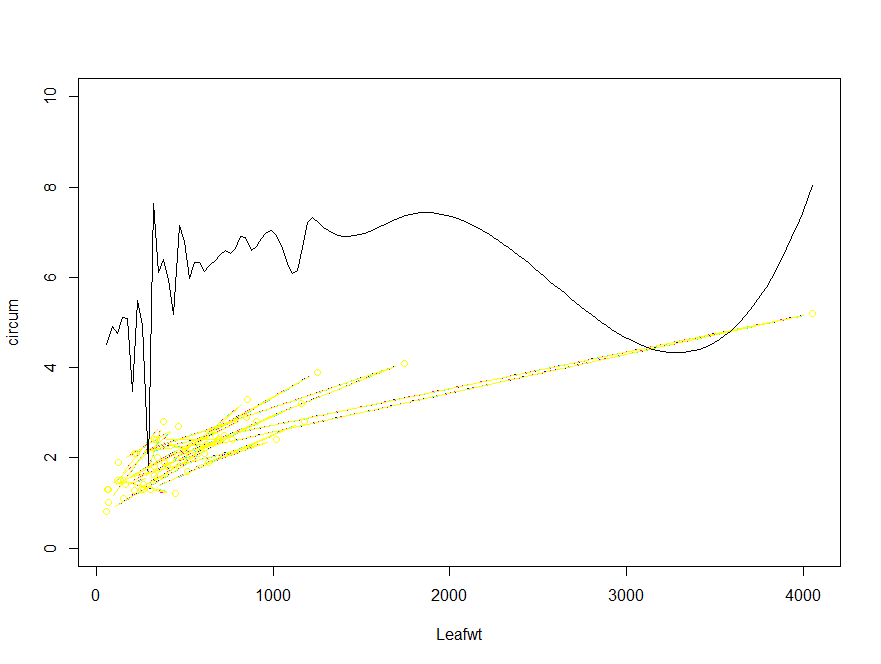
> for(i in 3:7)

+ {

+ lines(mesq[,8],mesq[,3],type="b",col=i,lty=i)

+ }

> lines(spline(LeafWt,pred.log.mesq))



(i)

> summary(log.mesq.lm)

Call:

lm(formula = log.LeafWt ~ Group + log.Diam1 + log.Diam2 + log.TotHt + log.CanHt + log.Dens)

Residuals:

Min 1Q Median 3Q Max

-0.70831 -0.12736 0.02225 0.13924 0.71277

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.7680 0.1551 30.747 < 2e-16 \*\*\*

GroupMCD 0.5834 0.1287 4.534 5.37e-05 \*\*\*

log.Diam1 0.3938 0.2820 1.397 0.170

log.Diam2 1.1512 0.2102 5.477 2.75e-06 \*\*\*

log.TotHt 0.3943 0.3129 1.260 0.215

log.CanHt 0.3732 0.2806 1.330 0.191

log.Dens 0.1093 0.1219 0.896 0.376

---

Signif. codes: 0 ?\*\*?0.001 ?\*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 0.3295 on 39 degrees of freedom

Multiple R-squared: 0.8873, Adjusted R-squared: 0.8699

F-statistic: 51.17 on 6 and 39 DF, p-value: < 2.2e-16

Remove "log.Dens":

> log.mesq.lm2=lm(log.LeafWt~Group+log.Diam1+log.Diam2+log.TotHt+log.CanHt)

> summary(log.mesq.lm2)

Call:

lm(formula = log.LeafWt ~ Group + log.Diam1 + log.Diam2 + log.TotHt +

log.CanHt)

Residuals:

Min 1Q Median 3Q Max

-0.75323 -0.15547 0.03622 0.15274 0.69783

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.7698 0.1547 30.836 < 2e-16 \*\*\*

GroupMCD 0.5382 0.1181 4.558 4.78e-05 \*\*\*

log.Diam1 0.4821 0.2635 1.830 0.0748 .

log.Diam2 1.1126 0.2052 5.422 3.07e-06 \*\*\*

log.TotHt 0.4240 0.3104 1.366 0.1795

log.CanHt 0.3846 0.2796 1.376 0.1766

---

Signif. codes: 0 ?\*\*?0.001 ?\*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 0.3287 on 40 degrees of freedom

Multiple R-squared: 0.885, Adjusted R-squared: 0.8706

F-statistic: 61.54 on 5 and 40 DF, p-value: < 2.2e-16

Remove "log.TotHt":

> log.mesq.lm3=lm(log.LeafWt~Group+log.Diam1+log.Diam2+log.CanHt)

> summary(log.mesq.lm3)

Call:

lm(formula = log.LeafWt ~ Group + log.Diam1 + log.Diam2 + log.CanHt)

Residuals:

Min 1Q Median 3Q Max

-0.75089 -0.15652 0.03434 0.15307 0.72143

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.8377 0.1480 32.685 < 2e-16 \*\*\*

GroupMCD 0.5691 0.1171 4.860 1.76e-05 \*\*\*

log.Diam1 0.5137 0.2653 1.937 0.0597 .

log.Diam2 1.1881 0.1997 5.949 5.12e-07 \*\*\*

log.CanHt 0.6658 0.1912 3.482 0.0012 \*\*

---

Signif. codes: 0 ?\*\*?0.001 ?\*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 0.3321 on 41 degrees of freedom

Multiple R-squared: 0.8796, Adjusted R-squared: 0.8678

F-statistic: 74.88 on 4 and 41 DF, p-value: < 2.2e-16

All the variables remaining are significant at 0.10, i.e., p-value < 0.10.

(j)

For initial model:

Multiple R-squared: 0.8873, Adjusted R-squared: 0.8699

For final model from the backward elimination:

Multiple R-squared: 0.8796, Adjusted R-squared: 0.8678

We can see that R2 decreased, since we have deleted two variables; However, the Adjusted R2 increased, which implies that those 2 deleted variables were not useful variables to the model (i.e., they improved the model by less than expected by chance), therefore I think it was ok to remove any variables that I removed.

(k)

> dgroup=as.numeric(Group)-1

> dgroup

[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> log.mesq.lm3=lm(log.LeafWt~Group+log.Diam1+log.Diam2+log.CanHt)

> summary(log.mesq.lm3)

Call:

lm(formula = log.LeafWt ~ Group + log.Diam1 + log.Diam2 + log.CanHt)

Residuals:

Min 1Q Median 3Q Max

-0.75089 -0.15652 0.03434 0.15307 0.72143

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.8377 0.1480 32.685 < 2e-16 \*\*\*

GroupMCD 0.5691 0.1171 4.860 1.76e-05 \*\*\*

log.Diam1 0.5137 0.2653 1.937 0.0597 .

log.Diam2 1.1881 0.1997 5.949 5.12e-07 \*\*\*

log.CanHt 0.6658 0.1912 3.482 0.0012 \*\*

---

Signif. codes: 0 ?\*\*?0.001 ?\*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 0.3321 on 41 degrees of freedom

Multiple R-squared: 0.8796, Adjusted R-squared: 0.8678

F-statistic: 74.88 on 4 and 41 DF, p-value: < 2.2e-16

> Groupn=(dgroup==0)

> Diam1n=myfun(1.7)

> Diam2n=myfun(1.5)

> CanHtn=myfun(1.4)

> als.mesq.pred=data.frame(dgroup=0,log.Diam1=Diam1n,log.Diam2=Diam2n,log.CanHt=CanHtn)

> pred.als.mesq=predict(log.mesq.lm3n,als.mesq.pred)

> pred.als.mesq

1

5.816075

> log.mesq.pred[30,]

Group log.Diam1 log.Diam2 log.TotHt log.CanHt log.Dens log.LeafWt

30 ALS 0.4700036 0.4700036 0.4700036 0.2623643 0 5.801816

> cbind(Diam1n,Diam2n,CanHtn,pred.als.mesq)

Diam1n Diam2n CanHtn pred.als.mesq

1 0.5306283 0.4054651 0.3364722 5.816075

> mesq[30,]

Obs Group Diam1 Diam2 TotHt CanHt Dens LeafWt

30 30 ALS 1.6 1.6 1.6 1.3 1 330.9

> exp(pred.als.mesq)

1

335.6521

Set "Group" as a dummy variable , then we get the predict value for "log.Leafwt" = 5.816075

The "log.Leafwt" of tree 30 is 5.801816 which is very close to the predict value for "log.Leafwt". The prediction of actual leaf weight is 335.6521 is also close to that of the tree 30, which is 330.9.

**Question 3:**

(a)

*CODE:*

>data tampa;

> infile '/home/shu.tu/tampa.txt';

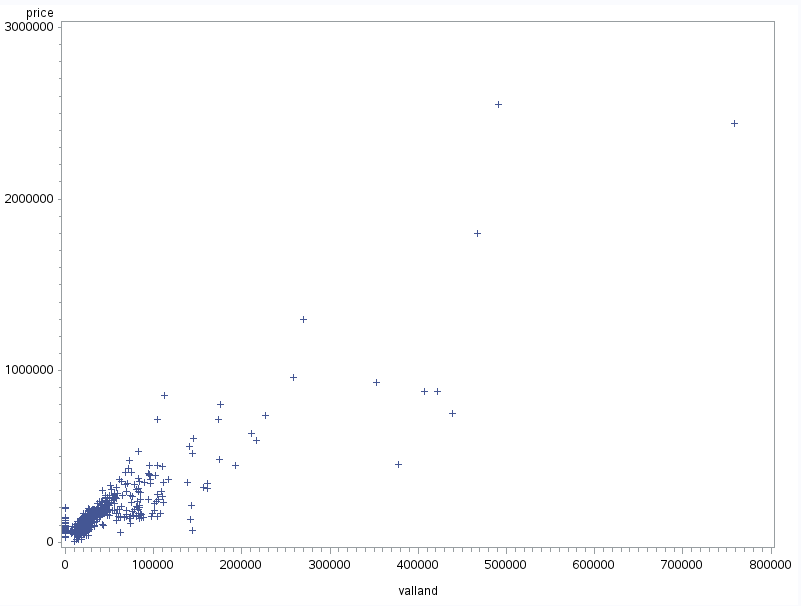
> input price valland valimp neigh$;

>

>proc gplot;

> plot price\*valland;

*RESULTS:*



(b)

*CODE:*

>data tampa2;

> set tampa;

> if valland >=150000 then delete;

>

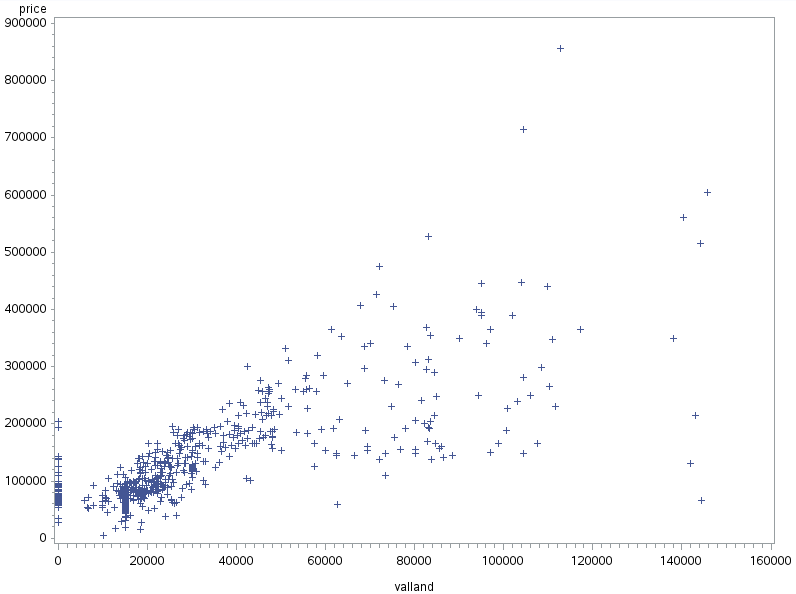
>proc gplot;

> plot price\*valland;

>

>run;

*RESULTS:*



(c)

The overall trend is a upward sloping line, I spot three groups of houses that are seriously out of line with the overall trend: i) a group is at the land value = $0, they are vertically distributed at zero; ii) a group whose land values are around $110000 and $ 120000 with selling price around $700000 and $850000 respectively; iii) a group whose land values are around $140000. On top of that, there are a bunch of houses whose land values between $60000 and $120000 are somewhat out of line, but not that seriously.

(d)

Not only there are a lot of outliers, but some of the outliers are influential observations that they will pull the regression function towards them.

(e)

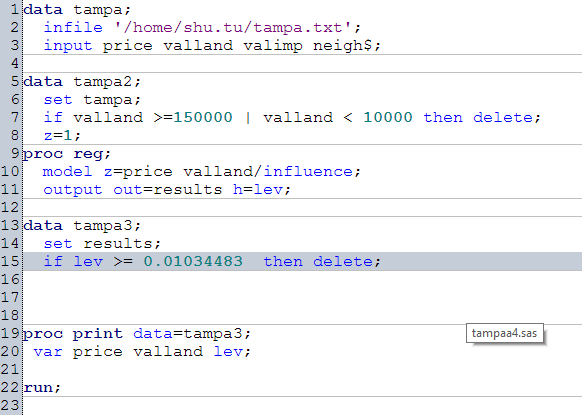
First, get rid of those land values smaller than $10000

Then, find out those very large and influential outliers and omit them.

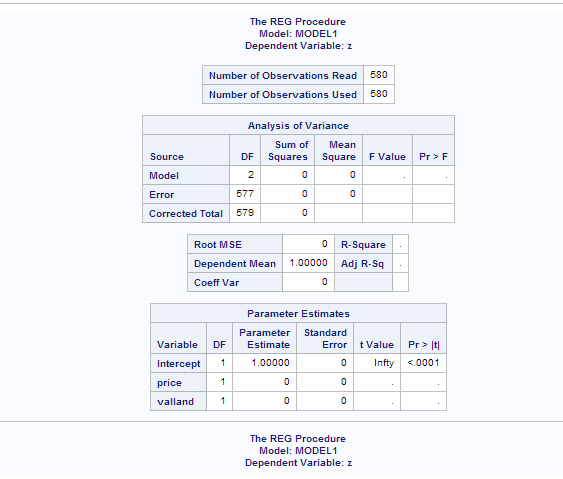
L = 2(k+1)/n = 2\*(2+1)/580 = 0.01034483, any hat diag h bigger than L is considered as influential outliers.

(f)

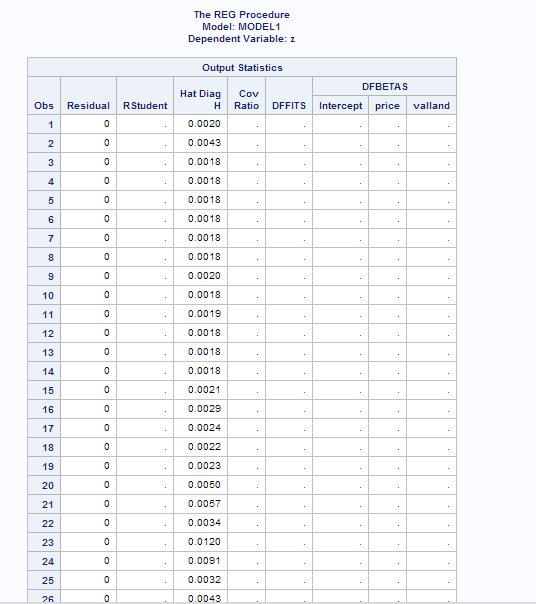
*CODE:*



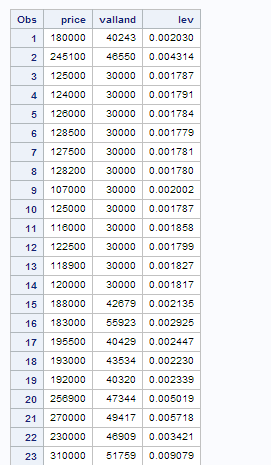
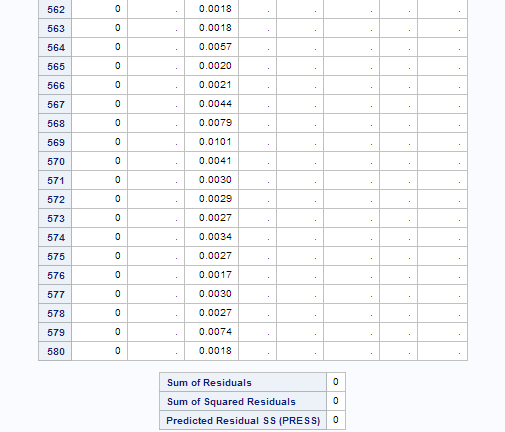
*RESULTS:*



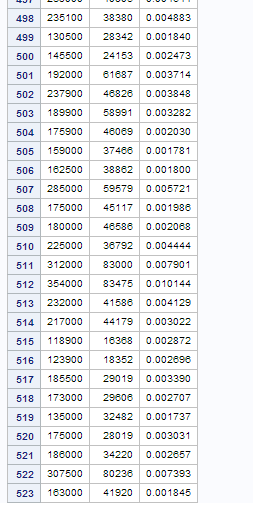
The data set "tampa2" which land valued between $10000 and $150000, has 580 observations.



**……**

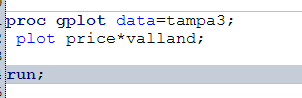


……

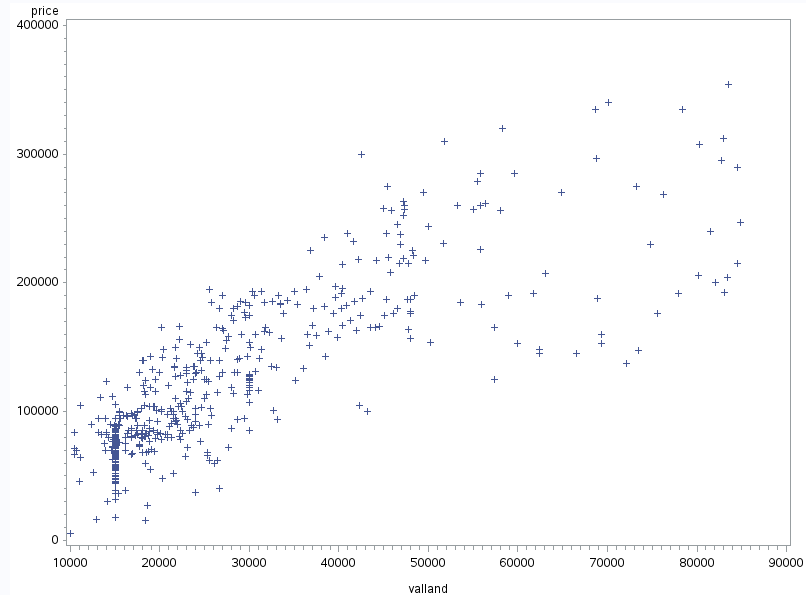


From the table we can see that after omitting the influential outliers from data set "tampa2", the new SAS data set "tampa3" has 523 observations.

*CODE:*



*RESULTS:*



From the plot we cannot find those influential outliers anymore.

**Question 4:**

(a)

> tampa=read.table("tampa.txt",header=F)

> colnames(tampa)=c("price","valland","valimp","neigh")

> head(tampa)

price valland valimp neigh

1 180000 40243 130189 TAMPALMS

2 245100 46550 166277 TAMPALMS

3 85400 100 68030 TAMPALMS

4 87900 100 65405 TAMPALMS

5 84200 100 68555 TAMPALMS

6 85000 100 64880 TAMPALMS

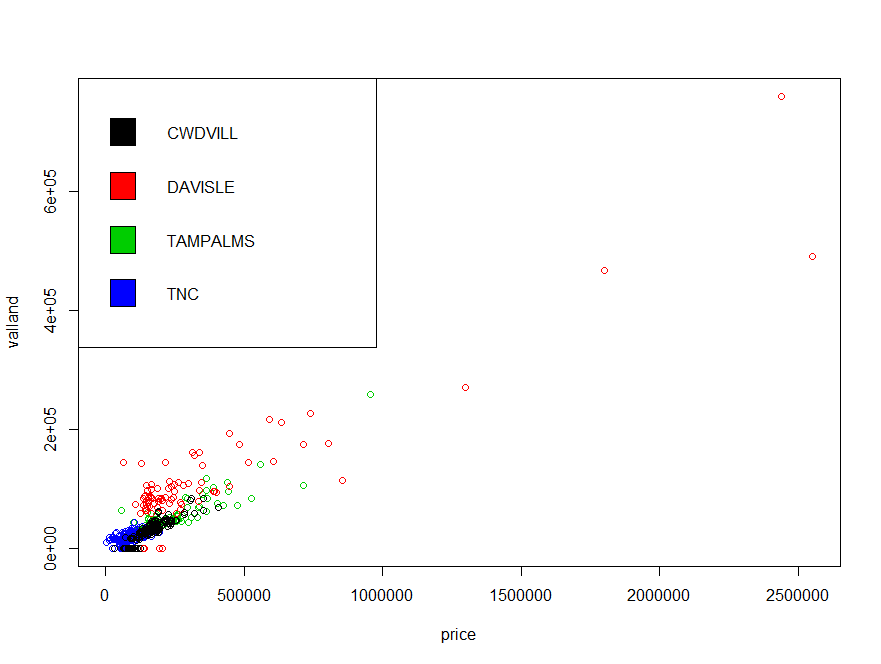
(b)

> attach(tampa)

> plot(price,valland,col=neigh)

> neighs=levels(neigh)

> legend("topleft",legend=neighs,fill=1:4)



(c)

Yes, from the plot we can see that "DAVISEL" has the largest spread(from 0 to 2500000) while "TNC" has the smallest spread(from 0 to 250000). And "TAMPALMS" has the second largest spread(from 0 to 1000000); "CWDCILL" ranges from 0 to 490000.

(d)

> tampa2=subset(tampa,select=c(price,valland,neigh),valland>10000)

> head(tampa2)

price valland neigh

1 180000 40243 TAMPALMS

2 245100 46550 TAMPALMS

8 125000 30000 TAMPALMS

9 124000 30000 TAMPALMS

10 126000 30000 TAMPALMS

11 128500 30000 TAMPALMS

> dim(tampa2)

[1] 600 3

(e)

> attach(tampa2)

> tampa3=cbind(tampa2,log(price),log(valland))

> head(tampa3)

price valland neigh log(price) log(valland)

1 180000 40243 TAMPALMS 12.10071 10.60269

2 245100 46550 TAMPALMS 12.40942 10.74828

8 125000 30000 TAMPALMS 11.73607 10.30895

9 124000 30000 TAMPALMS 11.72804 10.30895

10 126000 30000 TAMPALMS 11.74404 10.30895

11 128500 30000 TAMPALMS 11.76368 10.30895

> dim(tampa3)

[1] 600 5

(f)

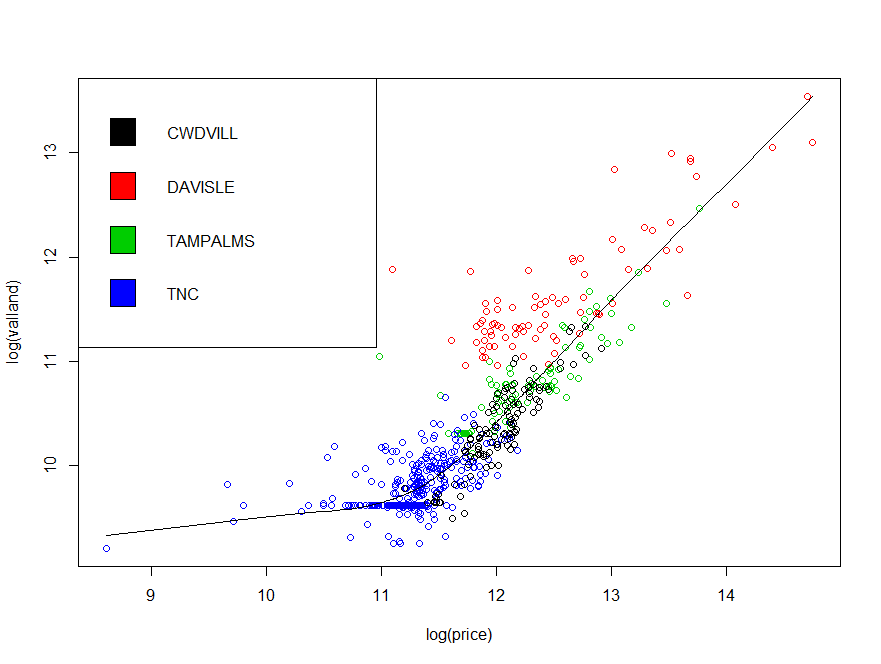
> attach(tampa3)

> plot(log(price),log(valland),col=neigh)

> neighs=levels(neigh)

> legend("topleft",legend=neighs,fill=1:4)

> lines(lowess(log(price),log(valland)))



(g)

Since the scatter plot is not linear, a linear regression would not be appropriate here.