

Course website	http://www.stat.wisc.edu/courses/st572-larget/
TA	Xu He(Mario)
Email	hexu@stat.wisc.edu
Website of TA	www.stat.wisc.edu/~hexu
Office	1275F, MSC
Office hour	Th 11:00-12:00, 4:00-6:00
Phone	(608)334-9792

1 Problem 1

Attention: All R codes used in my computer within discussion sections will be put on my website within 3 hours after discussions, from now on, if I used the projector.

Go back to the case of acorn size. Data available from <http://lib.stat.cmu.edu/DASL/Datafiles/Acorns.html>

Fifty species of oak trees grow in the United States. Twenty eight species of oak from the Atlantic region and 11 from the California region were studied. The size of each species' acorns was measured to see whether acorn size is related to geographic range.

In the first discussion, we have discovered that:

- a, Distribution of acorn size is not quite normal.
- b, There is somehow linear relationship between acorn size and tree height.
- c, Distributions of acorn size is somehow different in distinct regions.

Now, we will continue our study on the relationship between Acorn_size with regard to Range(quantitative) and Region(factor).

- 1, What is the mean and standard deviation of Acorn_size separately for each Region?
- 2, Plot the data to investigate the relationship between Acorn_size and Range. Plot the data by two groups with regard to Region.
- 3, Fit regression models with all Atlantic data. Make transformations if needed.
What is the most appropriate model for this data? How do you interpret this model?
- 4, Fit 3 regression models, using data in both regions and the same transformation used:
 - (1) Acorn_size with Range;
 - (2) Acorn_size with Range and Region, but without interaction;
 - (3) Acorn_size with Range and Region, and with interaction.
- 5, Of those 3 models in question 4, which one is the best model for this data? How do you interpret this model?
- 6, Use the final model from question 5 to predict the Acorn_size of acorns from California and with Range 1000.

1. Species: Latin name of the species
2. Region: Atlantic or California region
3. Range: The geographic area covered by the species in km²x100
4. Acorn_size: Acorn size (cm³)
5. Tree_height: Tree Height (m)

	Species	Region	Range	Acorn_size	Tree_Height
1	Quercus_alba_L.	Atlantic	24196	1.4	27.0
2	Quercus_bicolor_Willd.	Atlantic	7900	3.4	21.0
3	Quercus_macrocarpa_Michx.	Atlantic	23038	9.1	25.0
4	Quercus_prinoides_Willd.	Atlantic	17042	1.6	3.0
5	Quercus_Prinus_L.	Atlantic	7646	10.5	24.0
6	Quercus_stellata_Wang.	Atlantic	19938	2.5	17.0
7	Quercus_virginiana_Mill	Atlantic	7985	0.9	15.0
8	Quercus_Michauxii_Nutt.	Atlantic	8897	6.8	0.3
9	Quercus_lyrata_Walt.	Atlantic	8982	1.8	24.0
10	Quercus_Laceyi_Small.	Atlantic	233	0.3	11.0
11	Quercus_Chapmanii_Sarg.	Atlantic	1598	0.9	15.0
12	Quercus_Durandii_Buckl.	Atlantic	1745	0.8	23.0
13	Quercus_Muehlenbergii_Engelm	Atlantic	17042	2.0	24.0
14	Quercus_ilicifolia_Wang.	Atlantic	4082	1.1	3.0
15	Quercus_incana_Bartr.	Atlantic	3775	0.6	13.0
16	Quercus_falcata_Michx.	Atlantic	13688	1.8	30.0
17	Quercus_laevis_Walt.	Atlantic	3978	4.8	9.0
18	Quercus_laurifolia_Michx.	Atlantic	5328	1.1	27.0
19	Quercus_marilandica_Muenchh.	Atlantic	18480	3.6	9.0
20	Quercus_nigra_L.	Atlantic	10161	1.1	24.0
21	Quercus_palustris_Muenchh.	Atlantic	8643	1.1	23.0
22	Quercus_Phellos_L.	Atlantic	9920	3.6	27.0
23	Quercus_rubra_L.	Atlantic	28389	8.1	24.0
24	Quercus_velutina_Lam.	Atlantic	21067	3.6	23.0
25	Quercus_imbricaria_Michx.	Atlantic	14870	1.8	18.0
26	Quercus_myrtifolia_Willd.	Atlantic	2540	0.4	9.0
27	Quercus_texana_Buckl.	Atlantic	829	1.1	9.0
28	Quercus_coccinea_Muenchh.	Atlantic	8992	1.2	4.0
29	Quercus_Douglasii_Hook._Arn	California	559	4.1	18.0
30	Quercus_dumosa_Nutt.	California	433	1.6	6.0
31	Quercus_Engelmannii_Greene	California	259	2.0	17.0
32	Quercus_Garryana_Hook.	California	1061	5.5	20.0
33	Quercus_lobata_Nee	California	870	5.9	30.0
34	Quercus_agrifolia_Nee.	California	803	2.6	23.0
35	Quercus_Kelloggii_Newb.	California	826	6.0	26.0
36	Quercus_Wislizenii_A._DC.	California	699	1.0	21.0
37	Quercus_chrysolepis_Liebm.	California	690	17.1	15.0
38	Quercus_vaccinifolia_Engelm.	California	223	0.4	1.0
39	Quercus_tomentella_Engelm	California	13	7.1	18.0

Core R codes:

```
library('arm')
acorn=read.table('acorn1.txt',header=T)
acorn
str(acorn)

#1
attach(acorn)
sapply(split(Acorn_size,Region),mean)
sapply(split(Acorn_size,Region),sd)

#2
xyplot(Acorn_size~Range)
xyplot(Acorn_size~Range, groups=Region)
xyplot(Acorn_size~Range, groups=Region, pch=c(1,3), key=list(space='top', columns=2,
  text=list(levels(Region)), points=list(col=trellis.par.get('superpose.symbol')$col[1:2],
  pch=c(1,3)) ) )
xyplot(Acorn_size~Range, groups=Region, type=c('p','r'))

#3
detach(acorn)
lm1=lm(Acorn_size~Range, data=acorn[1:28,])
lm1=lm(Acorn_size~Range, data=acorn[Region=='Atlantic',])
display(lm1)
display(lm1,digit=5)

xyplot(log(Acorn_size)~Range, data=acorn[1:28,], type=c('p','r'))
lm2=lm(log(Acorn_size)~Range, data=acorn[1:28,])
display(lm2,digit=5)

xyplot(log(Acorn_size)~log(Range), data=acorn[1:28,], type=c('p','r'))
lm3=lm(log(Acorn_size)~log(Range), data=acorn[1:28,])
display(lm3)

#4
lm4=lm(log(Acorn_size)~log(Range), data=acorn)
lm5=lm(log(Acorn_size)~log(Range)+Region, data=acorn)
lm6=lm(log(Acorn_size)~log(Range)*Region, data=acorn)

display(lm4)
display(lm5)
display(lm6)

xyplot(log(Acorn_size)~log(Range), data=acorn, type=c('p','r'))
xyplot(log(Acorn_size)~log(Range), data=acorn, groups=Region, type=c('p','r'))

#5
```

```
xyplot(log(Acorn_size)~log(Range), data=acorn[-39:], groups=Region, type=c('p','r'))
lm7=lm(log(Acorn_size)~log(Range)*Region, data=acorn[-39:])
display(lm7)

#6
exp( lm7$coef[1] + lm7$coef[2] * log(1000) + lm7$coef[3] * 1 + lm7$coef[4] * log(1000))
exp(predict(lm7,data.frame(Region='California',Range=1000)))
xyplot(Acorn_size~Range, data=acorn[29:39:])
```