

HW7

Shaojie Zhang (*shaojiez*)

03/21/2019

Problem 1

Part a

```
library(MASS)
data(cats)

lm1 = lm(Hwt~0+Bwt:Sex, data = cats)
coef(lm1)
```

```
## Bwt:SexF Bwt:SexM
## 3.883453 3.914610
```

Forcing the intercept to be zero is good because it does not make sense for a cat to have heart weight > 0 and body weight 0.

Part b

```
model1 = lm(Hwt~0+Bwt, data = cats)
model2 = lm(Hwt~0+Bwt:Sex, data = cats)
anova(model1, model2)

## Analysis of Variance Table
##
## Model 1: Hwt ~ 0 + Bwt
## Model 2: Hwt ~ 0 + Bwt:Sex
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     143 300.09
## 2     142 299.90   1   0.19544 0.0925 0.7614
```

Comparing the two models using F test, we notice that the residual sum of squares are almost the same. And also the P value is quite large so we should accept the null hypothesis and reject the alternative hypothesis. In our case, the null hypothesis is that the coefficient for male is equal to the coefficient for female. So we don't have enough evidence to conclude that the coefficients are different.

Part c

```
Ttest = (model2$coefficients[1] - model2$coefficients[2])^2

that.star = vector(length = 1000)

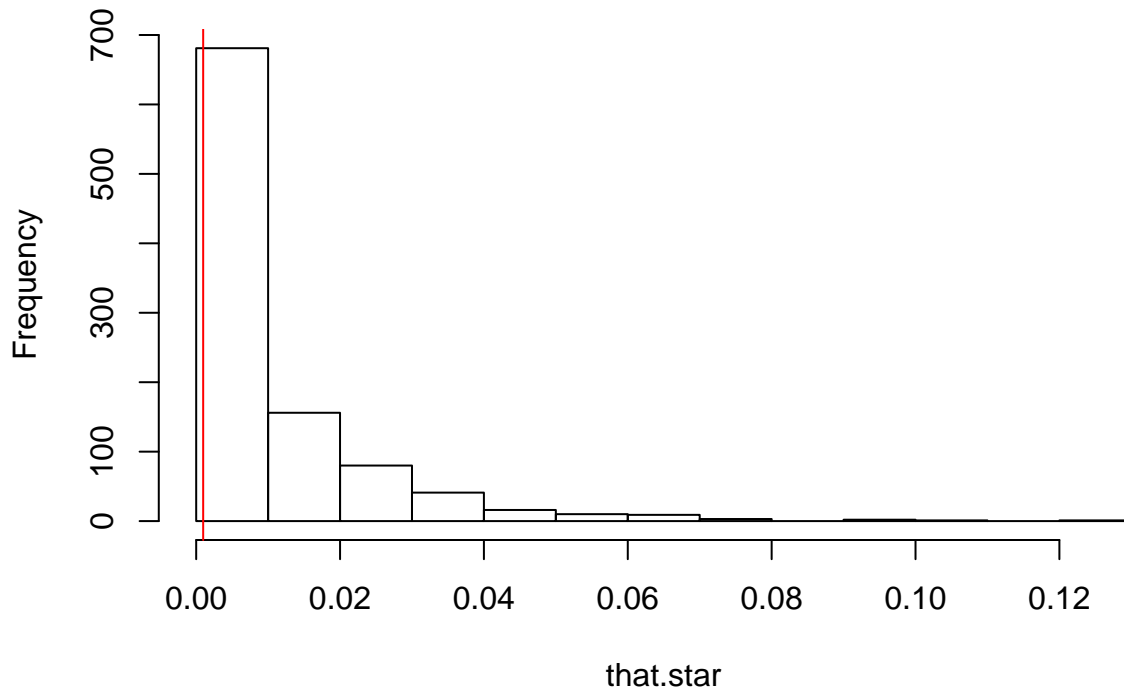
set.seed(1000)
for (i in 1:1000) {
  newSex = sample(cats$Sex, 144, replace = FALSE)
  newCat = data.frame(Hwt = cats$Hwt, Bwt = cats$Bwt, Sex = newSex)
```

```

modelnew = lm(Hwt~0+Bwt:Sex, data = newCat)
that.star[i] = (modelnew$coefficients[1] - modelnew$coefficients[2]) ^2
}
hist(that.star)
abline(v = Ttest, col = "red")

```

Histogram of that.star



```

pvalue=mean(that.star>=Ttest)
pvalue

```

```
## [1] 0.75
```

The permutation test is shown above. And we get a p value of 0.75, so we cannot reject the null.

Part d

Comparing the result with our previous result from HW6, here we have concluded that the coefficients for male and female might be the same. While in HW6 we found that the slopes differ. But note our variables are different. HW6 we did regression on sex only but now we are regressing on all kinds of interactions between sex. So the explanation of this effect is that maybe the interaction between male and HWT and female and HWT are the same.

Part e

If we look at the residual models, then they are different in the sense that here the distribution depends on Bwt*Sex while in the other homework it depended on only sex.

Problem 2

Part a

```
abalone=read.csv("abalonemt-1.csv",header=T)

abalone$Three=abalone$Length*abalone$Diameter*abalone$Height/(10^6)
model3fit=smooth.spline(abalone$Shucked.weight~abalone$Three, lambda = 0.0004)

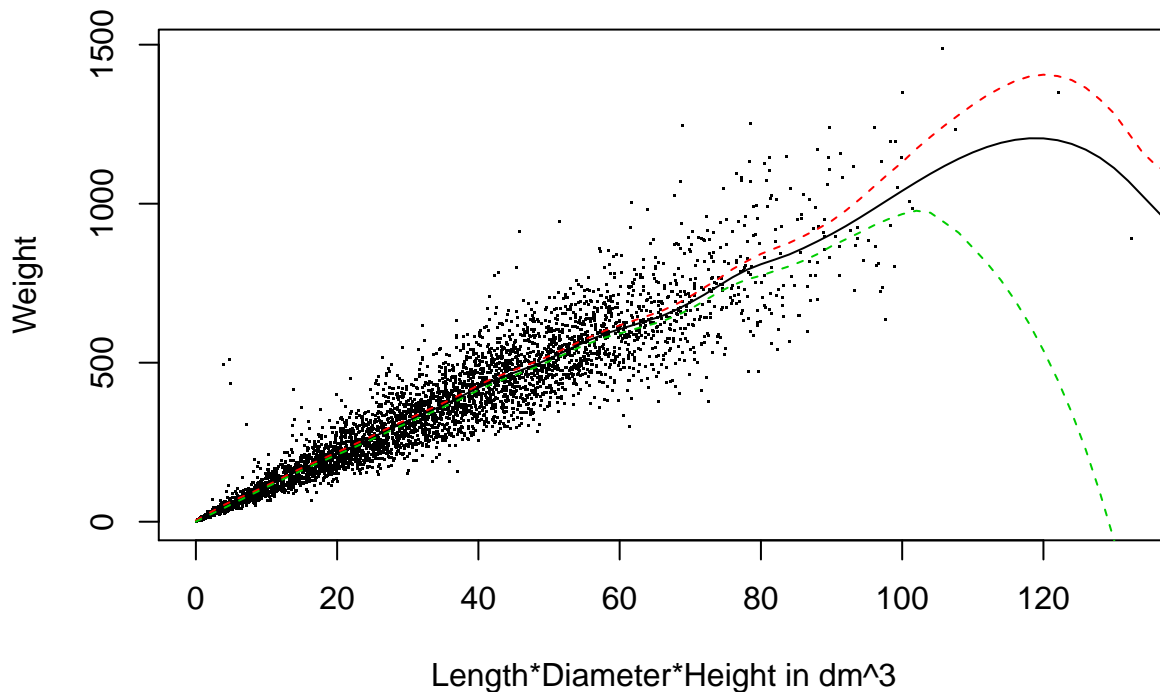
B=1000
x0=seq(0,140,2)
n=nrow(abalone)
prediction=NULL
samp=replicate(B,sample(1:n,n,replace=TRUE))

for (b in 1:B){
  temp=abalone[samp[,b],]
  ssmod=smooth.spline(temp$Shucked.weight~temp$Three,lambda=0.0004)
  pred=predict(ssmod,x0)$y
  prediction=rbind(prediction,pred)
}

quan=matrix(NA,nrow=2,ncol=length(x0))

for (i in 1:length(x0)){
  quan[,i]=quantile(prediction[,i],c(0.025,0.975))
}

pred1=predict(model3fit,x0)$y
plot(abalone$Shucked.weight~abalone$Three,pch=".",xlab="Length*Diameter*Height in dm^3",ylab="Weight")
lines(x0,pred1)
lines(x0,2*pred1-quan[1,],lty=2,col=2)
lines(x0,2*pred1-quan[2,],lty=2,col=3)
```



Part b

Note that the model is using different estimators. Our assumption is that the distribution of $\exp(\hat{r}(x)) - r(x)$ is similar to the one of $\hat{r}(x) * b - r(x)$

```
modellog=lm(log(Shucked.weight)~log(Length)+log(Diameter)+log(Height),data=abalone)
```

```
B=1000
```

```
x=data.frame(Length=c(600,400,250),Diameter=c(500,350,200),Height=c(150,125,140))
```

```
pred=NULL
```

```
for(b in 1:B){
```

```
  samp=sample(n,replace=T)
```

```
  temp=abalone[samp,]
```

```
  model=lm(log(Shucked.weight)~log(Length)+log(Diameter)+log(Height),data=temp)
```

```
  pred=rbind(pred,exp(predict(model,newdata=x)))
```

```
}
```

```
prediction=exp(predict(modellog,newdata=x))
```

```
apply(pred,2,mean)-prediction
```

```
##           1           2           3  
## -1.1258391 -0.8137884 -0.1275906
```

We see that the estimates of the bias are negative and gets closer. One of the possible reasons why this is happening is there are less variation.