Lab 8: Dummy variables and ANCOVA

Wayne Stewart

March 12, 2018

Use an interaction model to predict salary based on wage and gender

The data was taken from:

https://onlinecourses.science.psu.edu/stat502/node/188

https://onlinecourses.science.psu.edu/stat502/node/187

```
df = read.csv("salary-unequal.csv")
head(df)
```

```
##   gender salary years
## 1   Male     42     1
## 2   Male     62     2
## 3   Male     92     3
## 4   Male    112     4
## 5   Male    142     5
## 6 Female     80     5
```

```
df2 = read.csv("salary-equal.csv")
head(df2)
```

```
##   gender salary years
## 1   Male     78     3
## 2   Male     43     1
## 3   Male    103     5
## 4   Male     48     2
## 5   Male     80     4
## 6 Female     80     5
```

We can analyze the unequal slopes by adjusting and then using the following jags script

Task 1:

Remember to take the `eval=FALSE` off the chunk once you are ready. Please fill in the following Jags script and then run it.

```
require(rjags)               # Must have previously installed package rjags.
fileNameRoot="tut12" # For output file names.

#df = read.table(file="salary-unequal.txt", sep = "\t", header =TRUE)
df = read.csv("salary-unequal.csv")
df
salary = df$salary
```

```
gender = df$gender
years = df$years
GM = ifelse(gender == "Male", 1,0)




Ntotal = length(salary)  # Compute the total number of data rows
dataList = list(    # Put the information into a list.
  years = years,
  salary = salary ,
  GM = GM,
  Ntotal = Ntotal
)

#Define the model:
modelString = "
model{
for(i in 1:Ntotal)
{
mu[i]<- beta0 + beta1*years[i] + beta2*GM[i] + beta3*years[i]*GM[i]

}




}

" # close quote for modelString
writeLines( modelString , con="TEMPmodel.txt" )


#  initsList = list( theta=thetaInit )

initsList = list(beta0 = 0, beta1 = 0, beta2=0, beta3=0, sigma =10)

# Run the chains:
jagsModel = jags.model( file="TEMPmodel.txt" , data=dataList , inits=initsList ,
                        n.chains=3 , n.adapt=500 )
update( jagsModel , n.iter=500 )
codaSamples = coda.samples( jagsModel , variable.names=c("beta0", "beta1", "beta2" ,"beta3","sigma") ,
                            n.iter=33340 )
save( codaSamples , file=paste0(fileNameRoot,"Mcmc.Rdata") )

summary(codaSamples)

library(ggmcmc)
s = ggs(codaSamples)
ggs_density(s)

ggs_crosscorrelation(s)
```

Task 2:

Find parameter point and interval estimates. Interpret them!

Use these questions to help you in making interpretations in this lab:

1) If `x` is increased by one unit what happens to the mean value of `y`?
2) Will the lines ever intersect over the range of the data?
3) What meaning does the slope have?
4) You can answer Q.1 using point and interval estimates.
5) Make sure you know how to interpret BCI's (Bayesian credible intervals) – these are probability intervals and NOT Confidence Intervals.

Task 3:

Plot the data and the estimating lines

Taks 4:

Repeat task 1 - 3 for the equal slopes data and interpret the output. How should you change the model?