



**WORKSHOP  
DE BIOINFORMÁTICA  
APLICADA À GENÔMICA E  
MELHORAMENTO ANIMAL**



# **AULA PRÁTICA 4**

## **PLS em GWS**



**Fabyano Fonseca e Silva**

**Prof. Adjunto IV - Dep Zootecnia – UFV  
Estatística Genômica e Bioinformática**

**Campo Grande, 14/07 a 15/07 de 2014**

## PLS – Partial Least Square

```
gen=read.table("gs1_gen.txt",h=T) #lendo arquivo genotipos
dim(gen)
head(gen)

freq=read.table("gs1_alefreq.txt",h=T) #lendo arquivo freq alel
dim(freq)
head(freq)

fen=read.table("gs1_fen.txt",h=T) #lendo arquivo fenotipos
dim(fen)
head(fen)

#corrigindo fenótipo para efeitos fixos - de los Campos et al. (2013)

y1=mean(fen$y) + lm(y~ factor(ym_slg) + factor(farm) + hcw, data=fen)$residuals

library(pls)
fit_pls = plsr(y1 ~ M)

plot(MSEP(fit_pls)) #identificando o número ótimo de componentes

nc=10 #definindo número de componentes
snp_effect=fit_pls$coefficients[, ,nc] #obtendo vetor de efeitos de SNPs
snp_pls= cbind(colnames(M),snp_effect)
colnames(snp_pls)=c("SNP","a_pls")
write.table(snp_pls,"snp_pls.txt", row.names=FALSE,quote=FALSE)
plot(snp_effect)
```

```
gebv=M%*%snp_effect #obtendo vetor GEBV
rownames(gebv)=fen$id
plot(hist(gebv)) #distribuição dos GEBVs
top10=quantile(gebv[,1], probs = c(0.9)) #top10%
top10

# indivíduos selecionados e respectivos GEBV

sel_id=data.frame(gebv[,1][gebv[,1]>=top10])
colnames(sel_id)=c("GEBV")

#calculando h2 aproximada (geralmente subestimada)

va=var(gebv)
ve=sum((y1-gebv)^2)/(nrow(M)-nc)
h2 = va/(va + ve)
```

# #Método PLS: populações treinamento e validação

```
M1=M[1:500,]  
M2=M[501:634,]  
y11=y1[1:500]  
y12=y1[501:634]  
  
fit_pls1 = plsrf(y11 ~ M1)  
  
#identificando o número ótimo de componentes (treinamento)  
  
plot(MSEP(fit_pls1))  
nc=10 #definindo número de componentes  
  
snp_effect1=fit_pls1$coefficients[,nc] #vetor de efeitos de SNPs  
gebv2=M2*%snp_effect1 #obtendo gebv da pop de treinamento  
  
#acurácia  
  
cor(gebv2,y12)
```