HW6 BU CHONG

y<-matrix(c(19,29,24), ncol=1)  
n<-matrix(c(497,560,269), ncol=1)  
normal<-matrix(c(1,0,0), ncol=1)  
enlarged<-matrix(c(0,1,0), ncol=1)  
very<-matrix(c(0,0,1), ncol=1)  
data=as.data.frame(cbind(y,n,normal,enlarged,very))  
names(data)=c("yes","no","normal","enlarged","very")

result=glm(cbind(yes,no)~enlarged + very,family=binomial("logit"),data=data)  
result2=glm(cbind(yes,no)~very,family=binomial("logit"),data=data)  
m.step <- step(result, direction="backward",trace=TRUE)

summary(result2):  1.0297 labeled as residual deviance, i.e. delta G2 (the difference between the saturated model ie. ~enlarged+very and the current model i.e ~very???).

    Null deviance: 7.3209  on 2  degrees of freedom  
Residual deviance: 1.0297  on 1  degrees of freedom  
AIC: 19.881

Trace from stepwise:  1.0297 labeled as deviance, G2.  Based on a G2 of  1.0297 the cbind(yes, no) ~ very model fits well. This seems correct. The trace output make sense to me.

Start:  AIC=20.85  
cbind(yes, no) ~ enlarged + very

           Df Deviance    AIC  
- enlarged  1   1.0297 19.881 #~very  
<none>          0.0000 20.852 #saturated  
- very      1   7.2135 26.065 #~enlarged

Step:  AIC=19.88  
cbind(yes, no) ~ very

       Df Deviance    AIC  
<none>      1.0297 19.881  #~very   
- very  1   7.3209 24.172  #intercept only

> G2=1.0297  
> 1-pchisq(G2,df=1)  
[1] 0.3102291

Analysis of Deviance tables:  again 1.0297 is “residual deviance"???? Whereas deviance = 6.2912 for ~very, implying that the fit is bad for cbind(yes, no) ~ very???

> anova(result2)  
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(yes, no)

Terms added sequentially (first to last)

  Df Deviance Resid. Df Resid. Dev  
NULL                     2     7.3209  
very  1   6.2912         1     1.0297

Please explain this one too. Sorry to be so much trouble!!

> anova(result)  
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(yes, no)

Terms added sequentially (first to last)

         Df Deviance Resid. Df Resid. Dev  
NULL                         2     7.3209  
enlarged  1   0.1074         1     7.2135  
very      1   7.2135         0     0.0000

result3=glm(cbind(yes,no)~enlarged,family=binomial("logit"),data=data)

summary(result3)

Null deviance shows how well the intercept only model fits the data. Residual deviance takes care of the additional predictors. Adding VERY has decreased the value by (7.32 - 1.02 =) 6.3 on 1 df. This is a significant decrease in the likelihood, implying VERY is significant.

On the other hand residual deviance is 1.02 on 1 df. This is not significant, indicating that H0: Model with VERY fits.

In case of ANOVA, the same is seen. NULL indicates intercept only model and the value 6.29 is the difference of the two deviances.

The point to be careful about is that, when to test a more complex model againt a simpler model we look at the reduction in the deviance. If making a model simpler, i.e. getting rid of one or more parameters, does not increase values of deviance residuals, we prefer simpler model. In technical jargon: We do not reject the null hypothesis that the simpler model fits.

> y<-matrix(c(19,29,24), ncol=1)

> n<-matrix(c(497,560,269), ncol=1)

> normal<-matrix(c(1,0,0), ncol=1)

> enlarged<-matrix(c(0,1,0), ncol=1)

> very<-matrix(c(0,0,1), ncol=1)

> data=as.data.frame(cbind(y,n,normal,enlarged,very))

> names(data)=c("yes","no","normal","enlarged","very")

> data

yes no normal enlarged very

1 19 497 1 0 0

2 29 560 0 1 0

3 24 269 0 0 1

> result=glm(cbind(yes,no)~enlarged + very,family=binomial("logit"),data=data)

> result2=glm(cbind(yes,no)~very,family=binomial("logit"),data=data)

> m.step <- step(result, direction="backward",trace=TRUE)

Start: AIC=20.85

cbind(yes, no) ~ enlarged + very

Df Deviance AIC

- enlarged 1 1.0297 19.881

<none> 0.0000 20.852

- very 1 7.2135 26.065

Step: AIC=19.88

cbind(yes, no) ~ very

Df Deviance AIC

<none> 1.0297 19.881

- very 1 7.3209 24.172

> summary(result)

Call:

glm(formula = cbind(yes, no) ~ enlarged + very, family = binomial("logit"),

data = data)

Deviance Residuals:

[1] 0 0 0

Coefficients:

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

(Intercept) -3.2642 0.2338 -13.964 < 2e-16 \*\*\*

enlarged 0.3035 0.3015 1.007 0.31412

very 0.8475 0.3163 2.680 0.00737 \*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7.3209e+00 on 2 degrees of freedom

Residual deviance: 8.3267e-14 on 0 degrees of freedom

AIC: 20.851

Number of Fisher Scoring iterations: 3

> summary(result2)

Call:

glm(formula = cbind(yes, no) ~ very, family = binomial("logit"),

data = data)

Deviance Residuals:

1 2 3

-0.7565 0.6763 0.0000

Coefficients:

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

(Intercept) -3.0920 0.1476 -20.951 < 2e-16 \*\*\*

very 0.6753 0.2592 2.606 0.00916 \*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7.3209 on 2 degrees of freedom

Residual deviance: 1.0297 on 1 degrees of freedom

AIC: 19.881

Number of Fisher Scoring iterations: 4

> result3=glm(cbind(yes,no)~enlarged,family=binomial("logit"),data=data)

> summary(result3)

Call:

glm(formula = cbind(yes, no) ~ enlarged, family = binomial("logit"),

data = data)

Deviance Residuals:

1 2 3

-1.746 0.000 2.041

Coefficients:

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

(Intercept) -2.87998 0.15672 -18.377 <2e-16 \*\*\*

enlarged -0.08066 0.24664 -0.327 0.744

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7.3209 on 2 degrees of freedom

Residual deviance: 7.2135 on 1 degrees of freedom

AIC: 26.065

Number of Fisher Scoring iterations: 4

>