primary = read.csv(url("http://www.stat.ucla.edu/~cocteau/primaries.csv"),head=T)

names(primary)
dim(primary)

# transformations

primary$black06pct = primary$black06/primary$pop06
primary$hisp06pct = primary$hisp06/primary$pop06
primary$white06pct = primary$white06/primary$pop06
primary$growth = 100*(primary$pop06/primary$pop00 - 1)

# drop some counties
primary = subset(primary, state_postal!="MI" )
primary = subset(primary, state_postal!="FL" )
primary = subset(primary, !(state_postal=="WA" & racetype=="Primary") )
# Hosmer Lemeshow goodness of fit test

```r
fit = glm(winner~., data=primary3, family="binomial")
preds = predict(fit, type="response")
```

# 10 groups -- sometimes called deciles of risk

```r
br = cut(preds, quantile(preds, (0:11)/11))
```

# divide the predictions and the data into these groups

```r
grp_preds = split(preds, br)
grp_data = split(primary3$winner, br)
```

# form the mean in each group...

```r
n = sapply(grp_data, length)
exp_1 = sapply(grp_preds, sum)
exp_0 = n-exp_1
obs_1 = sapply(grp_data, sum)
obs_0 = n-obs_1
chis = sum((obs_1-exp_1)^2/exp_1)+sum((obs_0-exp_0)^2/exp_0)
1-pchisq(chis, df=8)
```

```r
plot(exp_1, obs_1)
abline(0,1)
```
# Hosmer Lemeshow goodness of fit test

```r
fit = glm(winner~region+Bush04,data=primary3,family="binomial")
preds = predict(fit,type="response")
```

# 10 groups -- sometimes called deciles of risk

```r
br = cut(preds,quantile(preds,(0:11)/11))
```

# divide the predictions and the data into these groups

```r
grp_preds = split(preds,br)
group_data = split(primary3$winner,br)
```

# form the mean in each group...

```r
n = sapply(grp_data,length)
exp_1 = sapply(grp_preds,sum)
exp_0 = n-exp_1
obs_1 = sapply(grp_data,sum)
obs_0 = n-obs_1

chis = sum((obs_1-exp_1)^2/exp_1)+sum((obs_0-exp_0)^2/exp_0)
1-pchisq(chis,df=8)
```

plot(exp_1,obs_1)
abline(0,1)
Inferring the action of a model from a table of regression coefficients can be a pain; what’s worse, it’s not really the size of the coefficients that matters, but instead the coefficient multiplied by the associated covariate.

In short, the “effect” of a covariate is more than just the coefficient; a **nomogram** is a graphical way to assess these effects; the term nomogram refers to a “graphical calculating device”
# import a cleaned up version of the primary data...
source("http://www.stat.ucla.edu/~cocteau/primary3.R")

# load a new library... this one has A LOT of functionality
library(Design)

# random stuff you have to do to get the nomogram to appear
dist = datadist(primary3)
options(datadist="dist")

# fit a logistic regression model with lrm rather than glm
fit = lrm(winner~.,data=primary3)

# make the nomogram! we have added a line at the bottom that
# is the probability...
invlogit = function(x) exp(x)/(1+exp(x))
nomogram(fit,fun=invlogit,funlabel="Probability Obama wins a county")
# Comparison with tables

# aggregate data into a table

```r
primary4 = aggregate(cbind("obama"=primary3$winner,"clinton"=1-primary3$winner),
                     list(Bush04=primary3$Bush04,region=primary3$region),sum)
```

```r
fit = glm(cbind(obama,clinton)~region+Bush04,data=primary4,family="binomial")
summary(fit)
```

```r
sum(residuals(fit,type="pearson")^2)
```

# chisq statistic

```r
n = primary4$obama + primary4$clinton
preds = predict(fit,type="response")
exp_1 = n*preds
exp_0 = n-exp_1
obs_1 = primary4$obama
obs_0 = primary4$clinton

chis = sum((obs_1-exp_1)^2/exp_1)+sum((obs_0-exp_0)^2/exp_0)
```

# unaggregated fit

```r
fit = glm(winner~region+Bush04,data=primary3,family="binomial")
summary(fit)
```