> setwd("D:/Documents (Louis Booth)/R/data")

> county <- read.csv("county\_data.csv")

> dim(county)

[1] 3220 13

> head(county)

State County TotalPop Men Women Hispanic White Black Native Asian Pacific IncomePerCap Unemployment

1 Alabama Autauga 55221 26745 28476 2.6 75.8 18.5 0.4 1.0 0 24974 7.6

2 Alabama Baldwin 195121 95314 99807 4.5 83.1 9.5 0.6 0.7 0 27317 7.5

3 Alabama Barbour 26932 14497 12435 4.6 46.2 46.7 0.2 0.4 0 16824 17.6

4 Alabama Bibb 22604 12073 10531 2.2 74.5 21.4 0.4 0.1 0 18431 8.3

5 Alabama Blount 57710 28512 29198 8.6 87.9 1.5 0.3 0.1 0 20532 7.7

6 Alabama Bullock 10678 5660 5018 4.4 22.2 70.7 1.2 0.2 0 17580 18.0

> county.1 <- county[,c(-5,-11)]

> county.1$dummy\_unemp <- as.numeric(county.1$Unemployment > mean(county.1$Unemployment))

> ###par(mfrow=c(3,3), mar=c(4,4,2,0.5))

> pdf(file="logit\_histograms.pdf")

>

> for (j in c(3,4,5,6,7,8,9,10)) {

+ plot(county.1[,12], county.1[,j], xlab="Unemployment",

+ main=paste("Boxplot of", colnames(county.1)[j]),

+ ylab="Count", col="lightblue")

+ }

> dev.off()

null device

1

>

> logistic.1 <- glm((dummy\_unemp=="1") ~ TotalPop + Men + Hispanic + White + Black + Native + Asian + IncomePerCap, data=county.1, family="binomial")

> coef(logistic.1)

(Intercept) TotalPop Men Hispanic White Black Native Asian

1.172200e+01 4.840685e-05 -9.530691e-05 -7.326829e-02 -7.232808e-02 -2.539476e-02 -5.099366e-02 -4.814129e-02

IncomePerCap

-2.300438e-04

> summary(logistic.1)

Call:

glm(formula = (dummy\_unemp == "1") ~ TotalPop + Men + Hispanic +

White + Black + Native + Asian + IncomePerCap, family = "binomial",

data = county.1)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.8647 -0.8313 -0.3935 0.9132 3.2961

Coefficients:

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

(Intercept) 1.172e+01 2.615e+00 4.483 7.35e-06 \*\*\*

TotalPop 4.841e-05 1.202e-05 4.026 5.66e-05 \*\*\*

Men -9.531e-05 2.441e-05 -3.905 9.43e-05 \*\*\*

Hispanic -7.327e-02 2.617e-02 -2.800 0.00511 \*\*

White -7.233e-02 2.641e-02 -2.739 0.00616 \*\*

Black -2.539e-02 2.660e-02 -0.955 0.33982

Native -5.099e-02 2.996e-02 -1.702 0.08874 .

Asian -4.814e-02 4.215e-02 -1.142 0.25335

IncomePerCap -2.300e-04 1.197e-05 -19.224 < 2e-16 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4423.1 on 3219 degrees of freedom

Residual deviance: 3324.5 on 3211 degrees of freedom

AIC: 3342.5

Number of Fisher Scoring iterations: 5

> p.hat.1 <- fitted(logistic.1)

> y.hat.1 <- round(p.hat.1)

> table(y.hat.1)

y.hat.1

0 1

1958 1262

> table(y.hat.1, y.true=county.1$dummy\_unemp)

y.true

y.hat.1 0 1

0 1470 488

1 321 941

>

>

> county.2 <- county.1[,c(-7, -8, -9)]

> logistic.2 <- glm((dummy\_unemp=="1") ~ TotalPop + Men + Hispanic + White + IncomePerCap, data=county.2, family="binomial")

> coef(logistic.2)

(Intercept) TotalPop Men Hispanic White IncomePerCap

8.886033e+00 5.090435e-05 -1.003481e-04 -4.461614e-02 -4.296265e-02 -2.325723e-04

> summary(logistic.2)

Call:

glm(formula = (dummy\_unemp == "1") ~ TotalPop + Men + Hispanic +

White + IncomePerCap, family = "binomial", data = county.2)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.8854 -0.8366 -0.3934 0.9180 3.0580

Coefficients:

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

(Intercept) 8.886e+00 4.058e-01 21.899 < 2e-16 \*\*\*

TotalPop 5.090e-05 1.171e-05 4.346 1.39e-05 \*\*\*

Men -1.003e-04 2.376e-05 -4.223 2.41e-05 \*\*\*

Hispanic -4.462e-02 4.207e-03 -10.606 < 2e-16 \*\*\*

White -4.296e-02 3.708e-03 -11.585 < 2e-16 \*\*\*

IncomePerCap -2.326e-04 1.147e-05 -20.280 < 2e-16 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4423.1 on 3219 degrees of freedom

Residual deviance: 3331.8 on 3214 degrees of freedom

AIC: 3343.8

Number of Fisher Scoring iterations: 5

> p.hat.2 <- fitted(logistic.2)

> y.hat.2 <- round(p.hat.2)

> table(y.hat.2)

y.hat.2

0 1

1960 1260

> table(y.hat.2, y.true=county.1$dummy\_unemp)

y.true

y.hat.2 0 1

0 1472 488

1 319 941