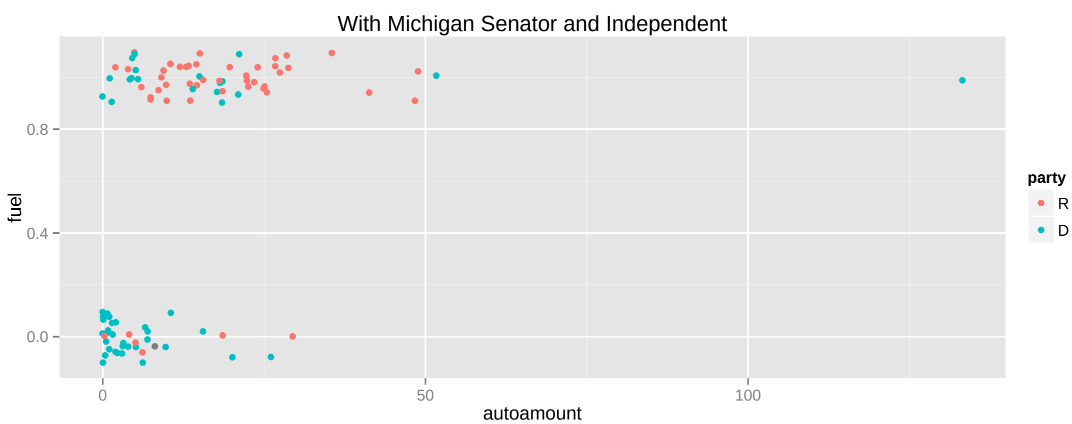
Stat 505 Assignment 9

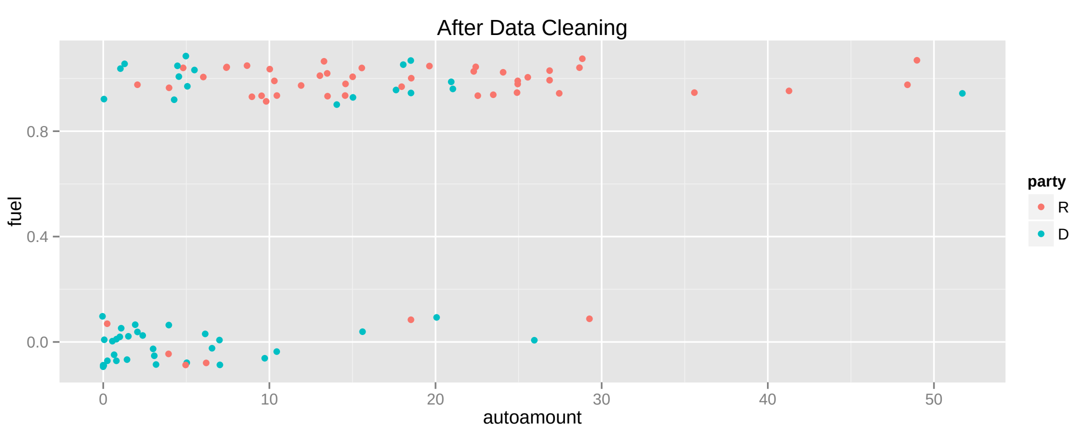
November 7th, 2014

1. For a doubling of the average price of cigarettes, the median quantity purchased in a US county is estimated to be associated with a multiplicative change of 1.23.

2. (a) The plot below shows the "YES" (coded as 1) and "NO" (coded as 0) votes across the amount of contribution from auto manufacturers, separated by party affiliation. Note that I changed auto contribution amount to \$1000s of dollars.

   First, I did some data cleaning. I threw out the one senator who affiliated with the independent party and as a result, I will restrict my inference to democratic and republican senators. There has also been some discussion about whether to include the Michigan senator that has a very large contribution from auto manufacturers. I ran the model with and without this senator, and the model did not change greatly. He does, however, cause problems with interpretation. We interpret the rate of change in the probability of a "yes" at the average value of the explanatory variable. The average contribution amount is greatly inflated if we leave in this Michigan senator. I will remove him and restrict my inference to those senators who received less than \$100000 from auto manufacturers.





(b) The output summary from the logistic regression model is shown below. The log odds of a republican senator voting "yes" when he/she receives \$0 from auto manufacturers estimated to be 0.477 with a 95% confidence interval from -2.203 to 1.495. After backtransforming, the probability of a republican senator voting "yes" when he/she receives \$0 from auto manufacturers is estimated to be 0.617 with a 95% confidence interval from 0.338 to 0.847.

Similarly, the probability of a democratic senator voting "yes" when he/she receives \$0 from auto manufacturers is estimated to be 0.204 with a 95% confidence interval from 0.096 to 0.357.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.4772	0.6000	0.80	0.4264
partyD	-1.8409	0.5739	-3.21	0.0013
autoamount	0.1083	0.0361	3.00	0.0027

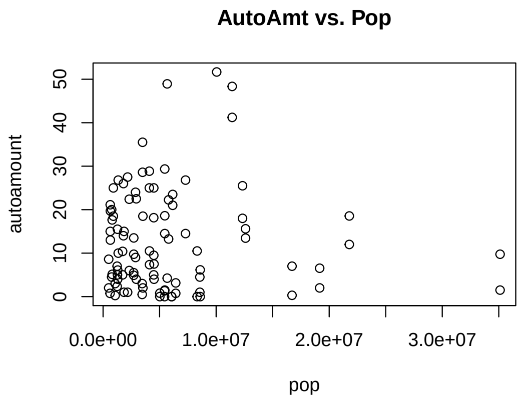
The log odds of a republican senator voting "yes" is estimated to increase by 0.511 for a \$1000 increase in contribution amount. Using Gelman's interpretation, an increase from $\$12,600$ to $\$13,600$ in contribution amount for a republican senator is estimated to be associated with a 0.0137 increase in the probability of the senator voting "yes", with a 95% confidence interval from a 0.0051 to a 0.0218 increase. Because we fit an additive model, the relationship for democratic senators is estimated to be the same.

(c) No, there is no evidence that a senator's vote depends on the population size of the state that they represent after accounting for party and contribution amount (p-value= 0.30866 from t-stat=-1.108 on 94 df). At a fixed contribution amount of 12.6 thousand dollars, the log odds of a senator voting 'yes' is estimated to decrease by 0.004491 for a \$100000 increase in state population size. Using Gelman's interpretation, an increase from 57.6 to 58.6 hundred thousand people is estimated to be associated with a 0.0005 decrease in the probability of the senator voting "yes", with a 95% confidence interval from a 0.0140 decrease to a 0.0035 increase.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6822	0.6414	1.06	0.2876
partyD	-1.7677	0.5805	-3.05	0.0023
autoamount	0.1078	0.0363	2.97	0.0030
I(pop/1e+05)	-0.0045	0.0044	-1.02	0.3087

I fit a linear model to evaluate whether the population size of states is related to the amount that auto manufacturers contribute to the state's senators. Based on the model I fit, there is no evidence the the mean amount contributed by automanufacturers to senators depends on the population size (p-value= 0.961 from t-stat= -0.049 on 96 df). I don't see a clear pattern in the scatterplot.

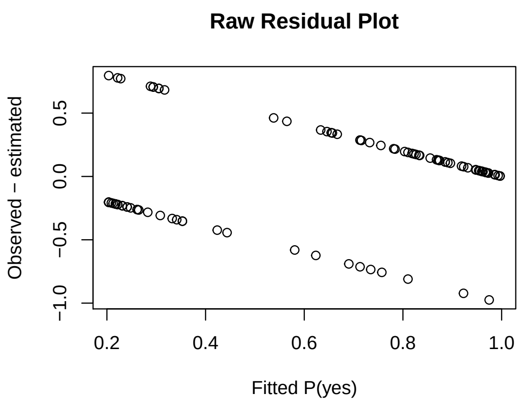
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.6991	1.5988	7.94	0.0000
pop	-0.0000	0.0000	-0.05	0.9612

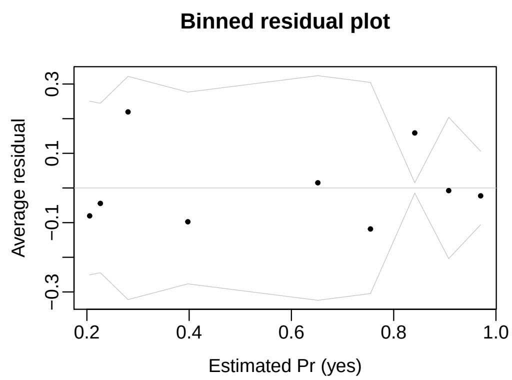


There is also no evidence that the the model with population fits better than the model without population (p-value= 0.2831 from drop-in-deviance chi-squared stat=1.1522 on 1 and 94 df).

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	95	88.36			
2	94	87.21	1	1.15	0.2831

(d) Below are binned residual plots for the model with party and contribution amount.





The binned residuals appear to be randomly scattered about 0, and we have only one residual that lies outside the grey ±2 standard error lines (we would expect to see either 0 or 1 of the nine residuals lie outside the gray lines). There is no indication that the model assumptions are violated and the model appears to fit well. There are only 98 points, however, which does not provide much information for the binned residual plot.

```
sen.dat <- read.table("~/Documents/Stat505/Homework/HW9/senators.txt", head=F)
names(sen.dat) <-c("last","first","state","party","fuel","autoamount")
```

```
sen.dat$autoamount <- sen.dat$autoamount/1000
sen.dat$party <- with(sen.dat,factor(party, levels=c("R","D")))
sen.dat$fuel<-with(sen.dat, ifelse(fuel=="YES",1,0))
require(ggplot2)
qplot(autoamount, fuel, data = sen.dat, geom="point", colour=party, position = position_jitter(w = 0.1, h = 0.1), main="With Mic
```

```
sen.dat<-subset(sen.dat,party!="I")
sen.dat<-subset(sen.dat,autoamount<100)
vote.fit1<-glm(fuel~party+autoamount,data=sen.dat,family=binomial)
qplot(autoamount, fuel, data = sen.dat, geom="point", colour=party, position = position_jitter(w = 0.1, h = 0.1), main="After Da
```

```
require(xtable)
xtable(summary(vote.fit1))
```

```
party.d <- with(sen.dat,factor(party, levels=c("D","R")))
vote.dem<-glm(fuel~party.d+autoamount,data=sen.dat,family=binomial)
require(arm)
invlogit(vote.fit1$coef)
invlogit(confint(vote.dem))
exp(vote.fit1$coef[1])/(1+exp(vote.fit1$coef[1]))
invlogit(confint(vote.fit1))
invlogit(vote.dem$coef)
sum(sen.dat$autoamount)/length(sen.dat$autoamount)
#12.6467
predict.glm(vote.fit1, data.frame(party="R", autoamount=12.6467))
#1.846531
coef(vote.fit1)[3]*exp(1.846531)/(1+exp(1.846531))^2
0.04373621*exp(1.846531)/(1+exp(1.846531))^2
0.1853789*exp(1.846531)/(1+exp(1.846531))^2
```

```
pop <- read.csv("~/Documents/Stat505/Homework/HW9/pop.csv",head=T)
pop <-pop[,1:2]
pop <- pop[-c(1,10),]
names(pop) <- c("state","twopop")
pop$state.abrev <- c("AL","AK","AZ","AR","CA","CO","CT","DE","FL","GA","HI","ID","IL","IN","IA","KS","KY","LA","ME","MD","MA","M
pop.alph <- pop[order(pop$state.abrev),]
pop.c <-pop.alph[rep(seq_len(nrow(pop)), each=2),]
pop.c <- pop.c[-c(44,92),]
sen.dat$pop<-pop.c[,2]
```

```
sen.dat$pop <- as.numeric(gsub(",","", sen.dat$pop))
vote.pop<-glm(fuel~party+autoamount+I(pop/100000),data=sen.dat,family=binomial)
xtable(summary(vote.pop))
```

```
require(arm)
invlogit(confint(vote.pop))
invlogit(vote.pop$coef)
sum(sen.dat$pop)/length(sen.dat$pop)
sum(sen.dat$autoamount)/length(sen.dat$autoamount)
#5764577
predict.glm(vote.pop, data.frame(party="R", pop=5764577, autoamount=12.6))
#1.781828
coef(vote.pop)[4]*exp(1.781828)/(1+exp(1.781828))^2
coef(vote.pop)[4]*exp(1.781828)/(1+exp(1.781828))^2
```

```
lm.check <- lm(autoamount~pop,data=sen.dat)
xtable(summary(lm.check))
plot(autoamount~pop,data=sen.dat, main="AutoAmt vs. Pop")
```

```
xtable(anova(vote.fit1,vote.pop, test="Chi"))
```

```
require(arm)
pred.8 <- fitted(vote.fit1)
br.8<-binned.resids(pred.8, sen.dat$fuel-pred.8,nclass=sqrt(98))
par(mfrow=c(1,2))
plot(pred.8, sen.dat$fuel-pred.8,ylab="Observed - estimated",xlab="Fitted P(yes)",main="Raw Residual Plot")
plot(range(br.8$binned[,1]),range(br.8$binned[,2], br.8$binned[,6],-br.8$binned[,6]),xlab="Estimated Pr (yes)",ylab="Average res
abline(0,0,col="gray",lwd=0.5)
lines(br.8$binned[,1], br.8$binned[,6],col="gray",lwd=0.5)
lines(br.8$binned[,1],-br.8$binned[,6],col="gray",lwd=0.5)
points(br.8$binned[,1], br.8$binned[,2],pch=19,cex=0.5)
```